Data Cleansing in Excel Made Easy
02/02/2023

CCH Learning:

Hello everybody and welcome to today's webinar, Data Cleansing in Excel Made Easy. My name is Susannah Gynther from Wolters Kluwer, CCH Learning and I will be your moderator for today. A few quick pointers before we get started. In the handout section, you'll find the PowerPoint slides for today's presentation. If you're having any sound problems, please check your settings in the audio section on the Go-to webinar panel, try to toggle between audio and phone. And just a reminder that within 24 to 48 hours, a notification for the e-learning recording will be emailed to you. You can ask questions at any point during the presentation by sending them through the questions box. I will collate those questions and ask them at the Q and A towards the end of today's presentation.

CCH Learning also offers a subscription service, which many people have termed Netflix for professionals. It provides members with access to our entire library of recordings as well as live webinars for a competitive flat fee. That's for over 500 hours of content. For CPD purposes, your viewing is logged automatically. Your presenter today is Ryan Racelis. Ryan is a certified Microsoft IT trainer and lends his expertise in the industry as an information technology consultant and trainer to various businesses ranging from preschool centers to business process outsourcing facilities. He also took the lead role in the development of business applications and productivity tools for various small and medium enterprise, both in the manufacturing and service sector. I will now pass you over to Ryan to begin today's presentation.

Ryan Racelis:

Thank you, Susannah. And just to make sure that you can hear me well and see my shared screen.

CCH Learning:

Yes we can.

Ryan Racelis:

Awesome. Good afternoon everyone. So in today's webinar we will try to cover the following, we'll be focusing on what you call Power Query. How can we use this to make the data cleansing in Excel a little bit easy or super easy? It depends on the level skill set that you have. So my focus in this one hour webinar is number one, is to identify how to use Power Query to connect your different data sources and we will go to the Power Query editor and let's learn on how we can use the buttons and the tools to do some cleansing, some scraping and reshaping your data so that the data will be ready for your reporting. And then at the very end we will learn how to load the data in your data model or a particular location on your Excel spreadsheet. So regarding about your data modelling though, if you have no clue about that, that's the Power Pivot.

I did a webinar of that one this morning. So check it out if you want to learn more about Power Pivot on the webinar this morning. Let me now go to this slide here. Our focus is to learn what you call Get and Transform. I will introduce to you Get and Transform, which is actually known as Power Query. Because when Power BI in Excel was introduced more than a decade ago, the term that the Microsoft use is what you called self-service business intelligence tools and Power Query is one of those tools. So back in the day they introduce Power Pivot,

Power Query, Power View and Power Map. So Power Pivot for data modelling, Power Query for your data transformation for the ETL process. And then we have Power Map and Power View for visualization. So regarding Get and Transform our Power Query, what we can do, number one, we can extract and discover your data source.

So what does it mean, your Excel workbook could have multiple tables. I can just grab a table or I can grab multiple tables within the workbook. Also, I can do what you call refining. I can't remove the columns that I don't need. I can remove the rows that I need. Sometimes the people who manages our data is giving us a hundred columns, but we only need three. We don't need those two million records, we just need those 10,000. So we can also use Power Query to combine our data sets. So first we will try to build some query and then we will also try to combine those query together to form a single source for us. So all of those will be discussed and shown to you in the demo in a bit. But before I go there, let me show you the different options on how to access Power Query depending on the version that you have right now.

So onto this slide, you will see from 2010 up to Office 365. So depending on your version, that is where you need to go to. So you need to look for Power Query. So if you have 2016 and then up to Office 365, you need to go to what you call the data tab because from 2016 up to Office 365 version Power Query is automatically embedded in your Excel application. Now on my next slide, depending on your version, just like what I said is you will see some windows or dialogue boxes that might be different. Okay, so let's now go to an Excel application. This is now the start of the demonstration. So first things first is again, if your version is not yet Office Excel 2016 or Office 2016 or Excel 2016 version, you need to add it in. So just to make sure it's clear, if you're using 2010 or 2013, you need to go to the file tab and then lower left you will see what you call the options.

And then you need to go to what you call add-ins. And on the manage dropdown you need to select COM add-ins, and when you click go, you will be able now to tick these boxes, excuse me. Then you need to look for Power Query and Power Pivot. Normally you tick those boxes, but as you can see on my setup, I don't need to do that because I'm using Office 365. It's already embedded here in this application. I can see that in the data tab. So somewhere here on the far left of my ribbon, I'm hovering my mouse right now. So ribbon on the far left. This is now the Get and Transform. So the first thing that I want us to know is what are the sources that I can use here in Power Query or Get and Transform? So first things first, we need to know what you call connectors. So Power Query use connectors to connect the different data sources. Now for us to see our option, we need to go to what you call the get data icon and let me hit that.

So there is a category of different connectors that I can use here in my Power Query. So the first one is what you call files. So I have here some Excel files, text or CSV, XML or JSON, PDF file or what you call folder. So what does it mean? Do you have some files in the folder that you want to use as a source a case of multiple files? So that can be used as a source using Power Query, okay. Then we have what you called a list of databases. So there are a bunch of this. Some people say, "Hey Ryan, what are those other databases?" So those are different database SQL platforms that you can use Excel to connect to as long as you're allowed to and grab some tables or views. So if your database, if you're familiar with database you're using and it's in the list, good on you, just ask the one who's doing the admin of your database to give you right to connect to the tables or views so that you can start grabbing the data rather than waiting for them to do it for you.

That's why back in the day it is coined as self-service business intelligence. As long as I'm allowed to connect to a database, I can just click this connector and I can use Power Query to grab the data and do some transformation. Now, something probably that you are using or not using it, we have what called some Azure sources. So it means these are sources from the cloud. We also have what's called power platform sources and we have online services. So majority of you, I strongly believe probably is using Dynamics 365 SharePoint Online and then Salesforce. So this are some online services that you can use Power Query to connect and grab your data sources. Then we also have other sources. So we can grab quickly from a table arrange where if there is a webpage wherein it contains the source that you want, there's a table structure that you want to grab.

You can also use Power Query to connect to that page. Then there is now a long list right here. There are some of them that is more on the database side, you can even write a query, so if ever that's the thing that you need to connect to. Okay, so there you go, I have given you here in Excel Power Query the different sources that you can connect to. Now for this demonstration, let me go to my file explorer. What we're going to do, number one is I want you to see the different sources that we will try to grab or to import in a single data model in our we'll push and load that to our Power Pivot. So our main file is staff information. So let me just quickly open this so that you know what is the content of this CSV file? So quickly on my screen I will quickly show you the structure, the problems that we have.

So for example, I have here a full name in to sign later on in the pivot table once to slice and dice or filter this based on the last name. So basically if you're an Excel user, you grab this source and then you split into columns. And you will also notice we will use this employee ID column because there is some key that we can use to connect to the other sources that we'll be grabbing. So regarding the department until the place of birth, we're happy with the data. These are categorical values, but look closely on the date of birth. Now on the date of birth column you will notice, okay, so it's using a post-op to separate those values. So obviously if you push this to Excel there will be problems because it will not be recognized as a date. So we will use Power Query to fix this.

Take note date of birth. The next one is H. This is a CSV file. So therefore this is a static value. So when you push this into your report, so this one will not change unless you change the source manually here. So we will try to replace this using Power Query to a more dynamic value. So employment date, current performance appraisal, we're happy with this one. Basic weekly salary, it's good to know everyone is getting paid. Look at the special allowance. So not everyone have an allowance. So we will also fix this column right here. So please take note it is open and Excel, this is why it's arranged this way, but this is a CSV file. Now I will not say that and then I will introduce to you to another source. So we have here what you call the calendar table.

So basically in the calendar table when it pops now to my screen, so you have a table. So we have the date, the year, quarter, the month, and the month ID right there. So that is now what you call a table. It's not a date in a spreadsheet, it is basically a table structure. You convert your range into a table by pressing what you call control T or the icon that I'm hovering my mouse right now on the upper left of my ribbon. Actually, we strongly suggest if you really want to use Excel as a source, so try your best to convert your range into a table. Also, in addition we have two more Excel files, so additional regional info and then NZ Regional details. So you'll notice when we grab this later on, so most especially the NZ Regional details, how bad the data is. Because our objective is use Power Query to make sure our data sources, our data set is arranged properly ready for reporting.

And lastly, I will give an example, how can I connect to a PDF file that contains a table structure. So this PDF file have multiple pages and one of the page contains a table structure and we will just grab the table so that we can relate that to our staff information, which is our main file. So you are a bit acquainted now with our data sources because of time, I just showed you the main ones and then we will just use Power Query to do the transformation. So let's go back now to Excel and just like we said, if we want to do some data cleansing, the best approach is Power Query. But before I start, probably some of you when you are doing some transformation is you're doing the old school, you're using the tools in Excel. So you're using copy and paste, you're using some function calculation, expression and then you will do a copy/paste value.

So that's fine, it will work. Some of you, because of skillset probably I'm now in the developer tab of the ribbon. If you put your attention on the far left right here, probably you have recorded the macro so that you can automate the way you want to do the data cleansing here in Excel. It'll also work or if you have a better skillset rather than recording, you will be writing your own macro from scratch. Now all of those personally I experience, I help some clients so it's all working. But one of the most things that I love that was introduced in Excel is Power Query because rather than me writing some scripts, writing some codes, doing it manually or recording a macro, there will be buttons that will be click or I can click so that transformation will happen to my data source. Now with that being said, let's grab one by one the different sources. So let's start a bunch of Excel spreadsheets. So let's go get data from file and from an Excel workbook.

Okay, so let's start here with the calendar table. So let's select the location and select the file and hit the import button. Now remember in this Excel workbook there is a sheet called calendar dates that contains some values, but that range of values was converted into the table. Now in my webinar this morning someone asked me a question, "Hey Ryan, is it prerequisite because I noticed when you're doing this demonstration you add the letter D, not the letter F, you add this, you add that." This is what you call naming convention. So because I've been working with database for more than two decades now, so that's the reason why there's some habits that a little bit hard to forget and get rid of. So those prefixes or naming convention help me easily to identify my dataset. So you can't ignore those prefixes. I can rename this calendar if I wanted to.

Okay, so in other words, you just need to make sure you select your source. Please take note. This one is the sheet. This one is the table. We strongly suggest to convert your range into a table because in Excel tables knows the beginning up until to the end. If you're selecting a sheet, the problem here, what if your colleague accidentally type something further, right or further down and then three months later you hit the refresh button. All of a sudden you will see what you call blank values because it will grab the whole sheet where you have used the cells placing some values. So my strong suggestion, convert your range in Excel to a table. Now let me click that, that's the one with the blue icon right here and let me click now Transform Data, it will lead me now to what you call the Power Query editor.

So there you go. I can see now what you call the Power Query editor. It's easy to manage and navigate this because if you're using Excel for quite some time, so you can see now the Power Query editor also have what call the ribbon. So at this point, so I tested it several times, there's a little bit of color right there, but it's called the home tab, transform tab, add column tab and the view tab. So there you go. I can see that I'm in a calendar table and this is the Power Query editor. This are the bunch of tools that I can use. Let me go navigate from one tab to another to transform this dataset. So on the left-hand side I'll see what you call the queries' pane. So later on, every time we grab more dataset sources, excuse me, you'll be able to see those different sources here on the queries pane.

Also on the right-hand side of my window, you will see what you call query settings. In the query settings you will notice we have the properties and the applied step. So I can use this property called name and then I can rename my query. Please take note, the output of your query is a table in your data model. Let me say that again. The output of your query will be a table in your data model, which is your power pivot. So even though the name in the source is the calendar, so if I don't want to use the prefix so that everyone understand what is this table all about, I can give it a better name so that everyone can understand. There is also a property that you can click here and then you'll be able to change some properties here. So add some description or take this new option called faster data loading. So let me hit cancel.

The next one, which is for me is one of the most important is what you called applied steps. So what happened was this. Remember we went to a source that is the first step. The second thing we did is we navigate and select that calendar table. This one that is actually the second step. And then after that we try to go here now to the Power Query editor. So the Power Query, try to detect what is the correct data, type in each column of your source. Now look at the first column. You can actually identify the data type. So on the left-hand side, when I click, I can see now and change from here the correct data type that I want, but I'm happy with dates. The year is a whole number. So that's one, two, three, and we can now go have a quick look on the other source.

Like the month, the month here ABC is actually a text data type. Now please take note, data type is different from formatting because if you are an Excel user and you don't have any clue about databases, so I just want to say the data type, I'm hovering my mouse right now in the ribbon. You can also change it from here on the left-hand side by the way is different from your formatting. Because in Excel we normally change formatting for the display purposes. Now in databases first before we can create the tables in the columns, we assign what data type it can accept. So that's how you build tables in databases. But in Excel, no, we can just come in, type the data and then we can now change the formatting. So the data type if I will try to compare it in Excel.

It is like a data validation rule. So again, to those who doesn't have any programming database background, who's attending today. So if you know Excel, probably you bump into a button or a tool that says okay, data validation. So you can create a rule in Excel and it'll only accept anything that is conforming to the rule. So somewhat the data type is something like that. So it means this column will only accept month ID, which is a number, which is a whole number. If I type here 1.1, it'll not be accepted. So here it will only accept the date. So if I type here my name Ryan, then it'll not be accepted because the data type will only accept date values. I hope I put some clarity. What do you mean by type, and what is your formatting context here in Power Query. Now, common question people keep on asking me, "Hey Ryan, where's the undo button? I'm an Excel user, so comfortable pressing control set and then undo so that I can go back to the previous step and then I can now replace that."

I just want to say there is no undo button here. So for example, if ever I accidentally remove a column, so I right-click month ID and I say remove, what happened, it added a step right here, so there's no undo button, but how can I bring it back? All you need to do is to just hit that X for you to remove that step and there you go. I can see again the month ID. Now please don't be quick and removing all of those steps because if you don't know how to bring it back, then you are in trouble because there's no undo button here. Also in my webinar this morning, someone asked a question, "Ryan, what if Susannah transferred this calendar source Excel spreadsheet in a different place in the network drive?" It's okay, as long as the structure is the same, you just need to know where it is located.

Now, what is the new name of the file? Because you can go back to the first step and then click this icon. So reference as a cog icon. So when I click that icon, I can now change the location. I can browse now somewhere in the network drive, click, click, click, click until I find the correct location in the new file name of my calendar. So grab it, hit import, and then basically if this is the same structure, it'll give you the same thing. And then try to go to the next step, making sure it's being applied properly and then until you reach the very end. So that's my strong suggestion. If the question is, "What if the source was changed in the file name or in the location?" So you need to go back in Power Query, go to the source step and then change and browse the new location or the new file name of your source.

Then when you're done, try to click one by one, making sure it's applied properly here in Power Query. Now the transformation that I want to add here is we have some calendar values like year, quarter, month ID and month. What I want to do is to quickly add what you call some financial year values. I want to add a financial year column and I want to add what you call the financial quarter column. So now those things, so can be done in different ways, let me say those things can be done in different ways. But if you'll be asking me if that's your question, I'll rather do it directly from the source. Let me say again, if those are the things that you want to do, I'll rather do it directly from the source. Because rather than doing it from here, so if that is something that someone asked me to do in the transformation here, so I'll rather do it in the source because I can quickly create that in the Excel spreadsheet.

Some organization, most especially big enterprises, big companies that they train, they put this in their SQL server or database somewhere so that everyone in the organization will just grab it quickly. But if ever you want to do something like that, just like we said, so there are many buttons that you can use. So you will notice if I select a date column and then I go now to the add column tab, I can see here a button called date and you will notice I can somewhat quickly grab some details regarding this date. But again, so if that's the case, I'd rather do it from the source. Another option rather than here, I will also do it in DAX. So I discussed a few DAX expression on the previous webinar this morning because DAX is more flexible compared to the buttons that you have here.

Now let's say I finish already the things that I want to do here. So because that's my objective, so I want to start simple. So let's say I did everything that I want. I'm happy with the state of my table right now, what do I need to do next? So you need to go to the home tab and then go to close and load and say close and load to. Now let me explain your option. When you say load to table, it will create a table within this spreadsheet or at this point it'll create a new worksheet or load to a pivot table report or a pivot chart. It means it'll create like a pivot table here and then the source is the one that you just transform using Power Query. But normally my suggestion to people is push this to your model.

So how to do that, only create a connection and then add this data to the data model. Because previously this morning I mentioned to you the data modeling in Excel can host more than two billion rows in each table. So that is amazing. We strongly suggest use this option, this combination, create a connection and add it to the data model. So when I hit the okay button, you will not see that here in this Excel spreadsheet, it will be hosted into your database engine that is the data model of this file. So how can I access that? I need to go to what you call the power pivot. If I want to view the data, the result of my Power Query, I cannot now go to the manage icon and it'll display it to me in the power pivot window. Now this one tells me, okay, these are the columns and then this are now the number of rows.

I'm now in record number one and I can add some calculated column to those who attend my webinar this morning. You saw some examples regarding that. So that is now your power pivot window. Okay, now let me go back to Excel and then go back to the data tab. So you might be wondering what if I want to refresh this data source? So simply you can just right-click and hit the refresh button. There's also a refresh button right there on the right and then there's also a refresh button right there in the ribbon. So there are many places for you to quickly refresh your dataset or your data source. So what does it mean? When I hit the refresh button it will follow all the steps that we just did in the Power Query editor in transforming your dataset, and then it'll update the dataset and obviously when the dataset approves your data model updates and then your pivot table updates basically. So it's like a quick ripple effect.

So in other words, that is pretty much amazing because I just hit one button and all those changes are happening. So that is actually my first part of my demo on how to grab from a source, how to be familiar with the window called Power Query, so the different components of the window and how to load it to the model. Now the next part of my demonstration is I will be a little bit faster. I will now go and grab those sources and then do now some transformation. So let's grab again another set of Excel spreadsheet. So let me now go to the Excel workbook here because we have first what's called additional regional info. So let me grab that and there is a sheet in the table I already explained convert to a table so that you can grab it easily and then if someone add a few more data further down or right. So it's okay, table knows the end from the beginning. So click the transform data because we need to do some transformation here. A quick one.

Now when it's loaded in my Power Query, number one, I can see now my second query, let me now give it a different name, a better name right there. So let me call this one region space NZ, because again it's a FYI, I'm training from here, Auckland, New Zealand. Okay, now there you go. I have some applied stuff right here, I don't need to discuss it anymore. So the thing here is I normally check if I'm happy first with the data type. So this are text. So happy with that one. I have a postal code, I have a latitude and longitude. By the way, the reason why I'm using this latitude longitude because I love displaying some reports using maps. Now here's the thing, if you can remember in our CSV file there's a first three characters that is attached in the employee ID.

So what do I need to do? I need to grab the first three characters here of the region. If you're an Excel person, you know how about left and right or mid-function, those are text function, but how can I do that here in Power Query? So please take note about this, focus on the upper left-hand side of my screen. We have a transformed tab and an add column tab. If you click a button in the add column tab, any button here, it'll automatically add a new column to your table. If you are in the transform tab, it will transform the selected column or the selected table, depending on which one that you're going to select here. So let me say again, transform tab, when you click a button or a tool, it'll transform a selected column or the selected table.

If you are in the ad column tab, when you push a button somewhere here or one of this, it'll add the new column for you. So let's say and assume I want to keep the region, I just want to create region ID, so I need to be in the add column tab. There's a button here, extract. So extract the first three characters because that's what I want to do. Remember three characters, I just type the number three and hit okay. So that is now an applied step right there and then I'll just give it a better name. So it's like Excel, I double-click and call it region ID. There you go. And I'm happy with this. One more time. I'm happy with this. Now if I know I have other sources that I want to push in here, what I can do is just go to the home tab one more time.

I will connect one more time in a source, another Excel file. So when I go to the home tab, if I go to the far right, I can see now what's called new source. So I will click this one and then there I go. I can see the list, remember I discuss it already. I just go and grab an Excel workbook again and then that is what you call the regional details right here. So I hit import and the one thing you'll notice here when it's now being displayed in this dialogue box, I don't have a table, I only have a sheet. Now just like we said, if you grabbing a sheet, you might have a problem, a worksheet, because what if there's no value in some of the rows in the column? So look at this one, this is a bad dataset, this is a bad source.

We need to transform this. We need it to be a better source. So let me select sheet one and push, okay, let's turn this into a better data source. So I have now the sheet one and then let me give this a different name. So let's call this one, so additional region details. It's a bit long name, but anyways, I just changed the name right there. Now the challenge is how can I turn it into a good source. So first things first, if you look closely enroll number one, when I click row number one, it gives me a list of values here in each of the column, it gives me now values. In Power BI, we are now in Power Query, there is a database, the engine that's being used in your model. So one more time, I use the word database.

In database, when you say now it means you don't know the value because in Excel, remember the first row here doesn't have a value. So when we try to import, it gives us what you call now value. Now, please take note when you say now in database, you don't know the value. It's different from zero. So if I don't need this first row, what I can do is I can just go here, click the dropdown and say I want to remove the top rows. See that? So in Excel, remember there's an icon on the upper left that you can click. Also, I can go to the ribbon and then look at this one, remove rows, I see that option. Remove top rows, those are same buttons, just in a different location. How many, just one. There you go, I'm happy with that. And then if you will look closely, the data is arranged in a horizontal manner.

Your database engine needs a table, a table structure is vertical. So how can I quickly change this? Let me go to transform, remember when I click something here, it'll change the column or transform the table. Look at this button here, I'm hovering my mouse to the upper left-hand side, transpose. See that? So it's like, "Okay, I'm almost done now." So one thing I know of is the first row now is not actually a record. The first row is actually a header. Now there's a button that says use the first row as header. Let me click that. There I go. I love now the structure of this table. Let me browse, making sure the data type is correct for me. There you go. I'm happy with the data type, as I browse. So there you go. I just click those button rather than me writing a macro or recording a macro or manually doing it in the source to tidy up so I can automatically use this if ever that's the case of my source.

Now again, I will say this, if you can do everything that you need in the source that is the best approach so that you don't need to do some steps like this. But we're using and we're discussing Power Query because based on my experience for my client here in New Zealand and also in Australia and also in Southeast Asia is your data is bad. So based on experience. So we need tools like this and other tools out there so that we can do the transformation and the reshaping of the data. That's why we're using and discussing Power Query. Now let's say I'm happy, let's go to the home tab and hit close and load to, making sure we're loading this to the model. Remember the combination, only create the connection and add this to the data model. So let's do that and you will notice it'll not be displayed in Excel spreadsheet.

It's now trying to load it into my model. I can see now those number of rows and records. How can I view it? I cannot now go to my power pivot. So let me now go to my power pivot. I can go here, it's in the middle option. So it says it's already here. There you go. So I can see now the region NZ and then also the additional regional details. There you go. This is now my data model. This is now my power pivot. So we have two sources left here. We have the PDF file and we have the CSV file. Let's grab that PDF file right there. Now I was happy when I saw that they finally use PDF as a source here. Now look at this one, I can go to get data from file and then grab the prompt PDF.

Remember this is your connector, it can import a data from a PDF document. Now you just need to make sure that the PDF document has a table structure. So hit import because let me show this one to you as I click from the option of what are the tables that I can grab? Now look at this. These are the different pages you will notice. So there you go. Those are the pages that they have in the PDF. But if I go to table one, Power Query recognize there was a table structure that you can grab here and then it gave it the name table 001 in page one. So let's transform this. Let's grab this and transform. Remember it'll lead me to the Power Query editor. There you go. I can see now it is page one. So the first row basically should be the header and we know that there's a button somewhere here. So I'm in the home tab somewhere in the center right.

Use the first row as header and then I can give it a better name. So employment types, so I have now those employment types. So this is the code Abbr, and this is the employment types column of this table employment types. Okay, so let us load this, make sure you load this again to your model, create only a connection. Now once this one is done, we're not viewing it anymore, because it's just a small table, we will now go to the last part of the demonstration. We're now focusing to our main file, the CSV file. So let's now go to get data and from a file we have CSV. We can also click that shortcut right there if this is basically... Because this one, these are some shortcuts that you can click, it'll work the same.

So the reason why I'm going here because this is the main location of your options, so let's grab that CSV file. So there's only one file in this folder. So let me grab the staff info, CSV. And this one is a little bit different because before it pushes it to your Power Query editor, we have now what you call the data type detection. Base on the first 200 rows of your data source, which is a CSV file, it went to each column and tried to identify what is the best data type in those respective column, based on the first 200 rows. By the way, we have 299 rows here. Excuse me. Now if ever your data source is heaps of rows or record.

For example, one time, around 2016, 17, I'm doing a data scientist professional program with Microsoft. I was finishing 10 papers so that Microsoft can call me like a data scientist. So I finish it in 2017. The reason why I'm telling that to you because one of our professor, so in one of the papers what the truth or is from NYU, and then she gave us some CSV file. And out of curiosity is when I grab the CSV file, now look at this, listen to this one, the CSV file she gave us 16 of those one CSV file is more than one gigabyte. Now hear that number, one gigabyte, so 16 different CSV file. Now out of curiosity, I try to use Power Query and then rather than basing on the first 200 rows, I select this one entire dataset.

I'm just trying to check how far this data type detection will go, now it's took me more than a day and then it's start cranking. So based on experience, if you have massive rows, you don't have to select the entire dataset. If you already know you have massive row, just stick with the first 200 row and I showed you already how to change the data type anyway. Now the thing that I love here is this column date of birth. Remember the date of birth, it's using a post-op to separate your details in each of those values. But right now this data type detection replace that with a forward slash so that it will be understood as a data type. That is amazing, by just looking at the first 200 rows. So it means I don't need to fix this, Power Query already fix it for me.

Now let's transform this data because we will be doing the last discussion about transformation, it'll lead me now to the Power Query editor. That's our main source right there, it's called the staff info. Now we will do this transformation. First, remember Susannah told me, "Ryan, I want to create the pivot table report, but I want to slice that using the last name. So how can we do that?" So my plan is to create a first name and last name in this data source. So how can I do that? Select the full name and I'll go to the add column. What I'm going to do, I will quickly duplicate the column, why I am duplicating the column? Because I want to give the full name and I want to have the first name and the last name. So what I can do now is in this duplicated column, I will go to transform because remember transform when you click a button or a tool here, it will transform the selected column or the selected table.

There's a button here called split column, and there you go. I have a bunch of option, but because I know my data, the full name contains only the first and the last name and separated by a space. Common question, people will tell me, "Ryan, what if only a few of them have a middle name, but Jordan doesn't have a middle name." If that's the case, you will keep adding some step until you reach that first name column, middle name column, and the last name column. It means I need to add more steps. So let me hit the delimiter and it'll tell me what's my plan on how to split this. So the delimiter, make sure it's a space and then I will use the leftmost delimiter in this instance and then one of the updates, for whatever reason, the default character is a code.

So we don't have that, so I will say none because there's no code that will be happening here. So this one, it's a good combination. This is just one of the updates. That's why I'm mentioning it. So let me hit, okay, because this column now will be split into two. And all I need to do is to give them a better name, now I can double-click, right click rename. So this will be the first name. Then last name or your surname, and there you go. So that's the first transformation that I want to add. Now let me go back here and I have now what's called the employee ID. Now remember I have employment types that I can use to connect here. Then I have also the regions that I can connect here, I used the region ID. So I cannot go to the employee ID right here, remember I want to grab the first three characters and also the last two characters.

So let me go to add column because I can go back to extract and say I want to extract the first character and add a new column. How many characters? Three. This will be my region ID. There you go, let me go back. Let me add one more. Let me grab the last two character. Make sure I'm in the add column tab because I want to add a new column for this. So two characters and let's call this one, employment types. Employment types, there you go. So I have the region ID, I can connect that to region NX. Employment types, I can connect here in the employment types table. So what else do we have to change here? So down with the full name, happy with the employee ID. There's nothing wrong based on our investigation regarding department up to place of birth. Then date of birth, I didn't do anything but Power Query fix it for me because of the data type detection, it's now giving me a date data type.

So that's now the icon for date data type. There you go. The age column, remember this is static, look at the data type ABC. ABC means it's a text value, so that's bad. So let's remove this. Now, remove column. Let's try to use Power Query to create a dynamic age for us here. So how can I add a new column for my age and I want it to be dynamic. Select the date of birth. Let me now go to the date icon again, I'm in the add column, please take note date, and I think you can see now what I'm going to click here. There is an age button. So if you attended my previous webinar this morning, I tried to showcase this one to you. There are many options here, but let me click now age. So I have 10 more minutes left here to finish this. It gives me a new column called age, and then it gives me the days difference between today's date and then the date of birth column.

So this number is the day's difference. I don't know about you when you're asked about your date, will you say, "Hey Ryan, I'm 12,433 days old." So probably we need to transform this. So now again, depending on your business process, it's up to you how, you just need to follow your steps in your process. Now let me transform this for us. So let me go transform tab and go to the far right we have here duration. Look at the duration. So heaps of option, because of time, let me move quickly, go to total years, and there you go. Now, if you want to truncate those decimal values, please take note, there are many options here to do this, but let me go to rounding. There's a roundup, round down or round, okay round depending on the decimal value that you want to remove.

Then roundup automatically if there's a value, it'll go round to the nearest zero right there, round down, it'll go down and truncate everything. So let me choose round down because I don't want to grow old so fast. So let me click round down. There you go, it just truncate. Now please take note, those three options will give you different result depending on the value that you have per record. I tried that several times, it depends on the value, it will give you different results. That's why you need to know what is your business process in calculating something. I'm happy now with this because this one is dynamic. So three months later or next year, it's okay, it'll automatically recalculate the age for me. I just need to hit the refresh button so that it'll grab now the new information for this, so data model. Employment date is all good.

Current performance appraisal rating, basically this area is amazing. That's fine. Special allowance, remember we have some blank and we know that now means unknown. So how can we replace this with the value? It's super simple. Let me click the first instance of that value now. There you go. You can see that here at the bottom in the transform there is a button that says replace values. When I click this, I can replace any value with another set of value. So I've used this many times in some of the clients, most especially with their phone numbers, so they have extension that they want to remove. They have an open close bracket they want to remove. So how can we do that here? Replace value when I push okay, all now will be now changed into zero. Make sense? Then, if I'm happy, I will now hit close and load because that's the last one we need to change.

This is now our fifth and this is our last caseload transformation that needs to be done. Remember when this is finished, the 299 rows will now look better. It'll be now good for your data modeling in your reporting. So let me now go to the data model. There you go. Now what is the next thing that needs to be done? Please go back to the webinar this morning, go back to the recording. I discuss it right there. How to build the relationship, I will not do it anymore here. So there will be now five sources, but this is the last thing for today. In the next five or so, six minutes. I will show you that you can combine two queries or sources. One more time. We will combine the two sources into one. So the objective is this, we'll be ending with this last part of the demonstration.

Let us go to the assumption that the only thing that we need in this particular column, which I'll give a wrong name, additional. So we just want to grab the island column. So we just want to grab the island column. Now what I'm trying to say, the region details was split into two tables. I just want to bring them together as a single table, but I don't want the other components or fields here. I just want to grab that island and I just want to push it here. The question is how can I do that? So let me answer that by showing it to you. First things first, I need to go back to the Power Query editor. So how can I do that if this is the case, for example, I cannot see a thing, so you need to be in the data tab.

So for example, I'm somewhere else, you need to be in the data tab. Then look for queries and connection because it'll give you this pane on the right so that you can now go to region NZ and edit the query. So right click and then hit the edit button. I gave some terms right there. I want to edit the query, "But Ryan, the output is a table." Yep, I gave the statement, the output of this query is a table in your data model. So I'm now here in the region NZ, my objective is to combine this with the region ID with the additional regional info. Now here is what I want to do. First things first, I want to create a region ID here because I noticed I don't have a region ID in this particular table. So remember, it's easy to do, from the add column tab I want to extract the first three characters because that's the one I'm going to use to link this two table. Let's call this one region ID.

Now I think I'm ready. So I have now the region ID, then I can now use this to merge it here on the region NZ. Now the end goal is I want to see the island column here only. I don't want to see all of the columns in that particular table. So how can I do that? First, select your main table or query and then go to the home tab. And there is something in the far right called the combined buttons. You need to select what you called merge query. Remember I want to merge the query together. Now select the column that you're going to use to connect or link the two tables. So there go region ID and then I will select now what you call the additional regional details. So, which I misspell, but it's okay, it's fine. People commit mistakes.

Let me select now the region ID right there because they're all the same. They have 16 rows for each table. I'm happy. And then hit the okay button, easy as that. And what happened here, Ryan? From my region NZ it added now a column for me. Please take note, if you know look up function. This is like a lookup function. Now let me explain this. Remember we have this NOR region ID. It's like a lookup function because when I go to this another column, it's like writing a lookup function. Can you see that? It's writing a lookup function. Now we're not done yet because I don't like the whole table, I just want the island column. So I will go to the right-hand side and there's an icon to split this table so I can select the column that I want. And then there are heaps of this.

So I will now quickly unselect all and just select the island and then don't use the original name as a prefix. When you untick this one, it'll say the column name will be island. If you tick this box, it'll say additional original details island. So tick this box and there you go. I have no island based on the region ID, but this island is based on the other query. This query right here, do not delete this query because something will go wrong in the region NZ. Now let me close and load because I'm changing it, now you will notice, so I will still have this query here. I will see now the five queries still. I'm changing something here, but I will do a little bit of a change in my power pivot in my data model. I should be able to see now island. There you go.

And what I'm going to do is I don't want this to be seen in my pivot table when I'm building a pivot table or a chart. So there's an option called hide from client tools. So it means I can see this in the model, I can see this in my query, but it'll not be displayed when I'm building or creating now some pivot tables, I'm hiding it. So now let me now go to pivot table. Now I'm creating a pivot table based on the model. You will notice I don't have that fifth table, I only have four. Then if I go to my queries... So pane, I still have five queries, but I have four tables only that I can use. So please take note, this one is not related yet. So if I want to basically create a good report here on using a pivot table, I need to make sure I relayed them properly.

So for me to quickly do that. So I'll go dates and then probably date of birth and then I'll go now to Abbr and employment types. So make sure I see everything here. So let me expand. So I know I'm going a bit out of time, so let me just go here. So to employment types and then region ID to region ID right there. And then bring it back to have the correct relationship, again, check the previous recording of the previous webinar this morning and then if I go back you will now see some change there. Then I can now create this pivot table here. So I can now grab something here. Let's start with the basic weekly salary and then I can now grab something from the island. Remember the island is coming from a different table, but what's pushed here, let me push that to the column and then I'll go to employment types to involve the other table and put it the row.

So this is now the employment [inaudible 01:01:49], something went wrong here because I see blanks, so it means I accidentally... Okay, so employment types here. So look at this one. The reason why I give blank, because I use, so are wrong, one, I grab types rather than Abbr. So how can I change this? So it's super simple, I can double-click this one. So as you notice, I can select the Abbr, so that I can see now, so that's the correct selection. Now I committed a mistake right there. I dragged the wrong field. Now let me have a quick check. There you go. Because if you have set up your model correctly, then the pivot table, there you go. That's what I'm expecting right there. So let me do some adjustment. And that's now my pivot table working, using now all of the three tables together and if I wanted to, I can even throw in the calendar. Now I know I went over time for two minutes, but I hope you'll learn something today with regards to Power Query. Back to you Susannah.

CCH Learning:

Thank you very much, Ryan. Appreciate all the information that you've been giving us today. We will be spending the next few minutes taking questions, so just a reminder to please type them into the questions' pane. To give you some time to type those up. I will mention our upcoming webinars. So coming up next week, we've got our first in our Cyber Security update series. We're also looking at Division 7A Essentials, Succession Planning for SMSF's and Choosing a Business Structure. We also have coming up Salary Packaging Opportunities and Family Law Financial Agreements, Back to Basics. If you can get details on all of these by heading to our website. So let's have a little look at our questions. Okay, so I have a question here, Ryan, from Michael. Michael is asking, now, I think you sort of answered this because he asked, "Would it be possible to use a PDF copy of a bank statement as a data source?" But then he also went in to ask, went on to ask, "Or a photographed one?" Say one that you've taken a photograph of something on a mobile phone.

Ryan Racelis:

Yep. Now let me answer that. That's a good question. Now at this point, the technology is something that they can, if the Power Query can distinguish, it's a table structure. Now, if it is a picture, I can tell, was 90 to 99% it might not give you the table, but from a bank statement PDF file, I will assume it is a table structure. The answer is yes. But regarding on the picture, I think the technology at this point is not there yet. But it's not part of our discussion, there's another tool to do that. Grabbing that picture and turning it into a table, but there's a tool, so it's called the Power Automate. So there's a tool like that they can use.

CCH Learning:

Okay, thank you for that, Ryan. So there you go, Michael. So we can do the bank statement, but possibly not the photograph without extra applications. I also have a question from Sarah. Sarah was asking, "Can multiple data sources be combined into a single source?"

Ryan Racelis:

Yes. That's actually the last thing that I did. Can I share my screen? I'll make sure, I will show that.

CCH Learning:

Pass it back to you.

Ryan Racelis:

Yeah. Let me select my screen. Make sure can see it now.

CCH Learning:

Yes.

Ryan Racelis:

Cool. Yep, okay. Now the only demo that I did, again, because of time. So it's what called merging. Now because of that question, it's not just merging, you can also do what you called append. Now the merging is the one that we just did. So if you can remember, I merged two queries. So what does it mean? I'm merging the columns together. There's another option, thank you for that one, because there is also what you call append queries, but this type of combining queries is your appending records. So for example, I help some government organization here in New Zealand, because of their system, they can only grab a portion of the data. So let's say this is the first set of data, the second and the third. So what I need to do at the very end is I go to the main table and then I will append those three queries into a single query. So regarding on that question, can we have that one? Yes, actually you have two options merging the columns or appending the records.

CCH Learning:

Thank you for that, Ryan. So I hope that helps you out there, Sarah. Well, that seems to bring us to the end of our questions for today, unless someone else out there has any other questions. No, that's okay. So in terms of next steps, I'd like to remind you all to please take a moment to provide your feedback when exiting. We have asked you a couple of questions about today's webinar, so it's really important for us to hear your opinions. It's also a reminder that within 24 to 48 hours you will be enrolled into the e-learning recording. Which can be watched multiple times and have access to the PowerPoint transcript, any other supporting documentation, and of course, a CPD certificate.

As mentioned in our previous webinar, we are having a bit of a problem with our notifications at the moment. So please head to your learner dashboard to find your e-learning recording if you don't get an email in regards to the e-learning recording. I would very much like to thank Ryan for the session today. Lots of information. So thank you very much and a big thank you to you, the audience for joining us. We do hope to see you back online for another CCH Learning webinar very soon. Enjoy the rest of your day, thank you very much.