2月4日datascience_tutorial_Tableau

Thu, 2/4 1:05PM 58:28

SUMMARY KEYWORDS

tableau, calculation, challenges, detail, tutorial, region, find, car park, data, showing, questions, date, expenditure, average, screen, return, co2 emissions, ruben, session, level

00:00

Law is the world. Thanks for joining up with today's lab. I'm just going to share a screen to remind you what I'd like you to

00:16

do. And, as, as you are doing and as you previously do

00:23

do pose questions in the chat, I will answer the initial questions when I'm just showing people what we're doing.

00:49

So, you should be able to stay, see the live screen, and it's pointing at the practicals area in the nervous for the module.

01:14

And there are

01:18

three things that I'd like to focus on today both to do a tableau but the first of those is to cover the material about calculations in the tutorial I wrote, and I have actually added some extra content to that. Today, which is useful to know and it will also be very useful to know on the challenges themselves. It is start by looking at the tutorial again, and specifically at sections 4.4. If I go to that. And then load the tutorial. Except I'm going to show you on my word version. You will say that there are no. I just want to zoom in and the original calculations, part of the tutorial. There was a section 4.4 point one and 454 point two those those have not changed. But I had added in these next two, four by four by three, and 4.4 point four, I'll just go to the first one of those zoom out inside of it. So, this makes use of the world indicators dataset the same one as you were using for the earlier parts of the calculations tutorial and finding null values. I'll let you read it and walk your way through it, but it should it introduces you to some very useful functionality, called if null, which you can use to actually find out about null values. And that's how we've ended up with these minus ones here and the columns for 2011 2012. So, this is, this is new today, I added it to work through that, that was useful. And then also new today is the calculations about level of detail level of detail is quite a complex concept, while the concept complex but it's the. It takes a bit of effort to get your head around again to work through that one there are a number of steps on it. This is here we got this figure, 23. An example of the level of detail calculation, which we end up using to show the Lv expenditure, as a, as a ratio of what is the average in a given region. And in this one you can see I just chose Middle East because there aren't very many countries in the Middle East, then some other areas. and you'll see that some countries, spend many many variable times per person on health, and other countries and the same would be true for any reason in the world. So, start off this tutorial by looking at that 4.4 point three and 4.4 point four. And I will, if you have quick queries, then post them on chat. If you and I will go briefly go through them. Well let's say about maybe the first one in about 10 minutes and then the second one or five minutes after that. Once you've done that, then please go on to the tableau challenges, and a challenges in there but the ones that I particularly want you to focus on the challenges, two to four. You probably won't get through all of those today but you will find that those calculations to do with now and level of detail are useful to you. Hopefully we can cover at least the first one in.

06:14

In this session. So, and just to answer some of the questions that Donna about recordings, and I have, I am talking to it to try to find out where the recording my recordings from the previous sessions are. I've got something like 50 recordings in my university zoom account, but none of them have for the sessions that I actually gave. So I'm not sure what's going

on there I don't yet have an answer from university it. And in terms of Ruben which questions are you talking about, okay, you've been challenges two to four. Download challenges yes I absolutely. And if you start by looking at those two new parts of the tutorial. Actually, They will then help you with the challenges. That's why I wrote them this morning that go back to those two new parts of the tutorial. You might then be able to work out how to do those challenges two to four. But I certainly hope to spend a bit of time on those showing you what the answers are during this session. I'll just stop sharing the screen and until I need to show you something. Keep the chat and the query is going either talk in chat or unmute yourself and go.

08:27

gallon the document I wrote this morning is the tableau tutorial on the practicals I have replaced the copy that was there with a new one. So, you go to practicals to the tableau getting started in tutorial, and click on the tutorial. All questions are good questions. Also email about the, the recordings from the sessions. That's what I explained a few minutes ago, I am talking to it to try to find out where they are. I have about 50 recordings in my university zoom account, but none of them are of the actual sessions that I ran down are regarding tableau, the place to start is where the getting started, and tutorial and start with the getting started. Part of that, which gives you hold your hand as you do some. If you create some simple visualizations from some of their sample datasets shall have what do you do now. I'm not sure if you've only just joined the session. If you don't come in on the practicals. Go to the Getting Started Tableau Getting Started tutorial. I have uploaded a new that I should have that tutorial today. And in that new version sections of five demand yeah four by 4.3 and 4.4 point four do those first that will. Okay, great.

13:16

do you to comment about somebody having the old version of the tutorial and are the students able to access the new version with these two sections, I'll share a screen again so it's absolutely clear. Okay, so I've got the new one. So I don't know try, I mean try simple, I mean simple like refreshing the screen whoever gf is referring to reload the page. I'm going to share our screen and step through the first one of those four by four by three.

15:43

Oh, that's a good idea Basheer. Share it that way. So just to be clear what I'm doing here. This is, this part of the tutorial the first new bit about finding null values. And they step through it bit by bit on there's got to recreate this job here. And in that that you will see

that there are power tells you that there are no values.

16:57

I'm changing the year to being continuous. I'm changing co2 being the average. co2. And you see down there in the bottom right hand, part of the screen. That is telling me there are 12 null values and that's what we want to investigate.

17:41

If we can work out. Well, when I know for a fact that it's to do with the co2 emissions that I can prove that by taking away the other pills and you're seeing the number of null values, it doesn't it doesn't change here because this is just like an average. Going back that we want to find out about these null values. And what you'll see. I'm doing in this calculation is setting $1 value to minus one. Now, you would expect that co2 emissions would always be positive. But we don't actually know that. And so before setting it to minus one, it is important to actually check that our assumption is correct. The simple way to do that by getting rid of these is simply to find out what is the minimum value of co2 emissions and 1000s of seven. So we've confirmed our assumption that co2 emissions values when they are provided are always positive. So what we're going to do is to use this if null function. And what this function does is if a given value is null, it will set it to whatever the second parameter is so we'll set it to minus one. And I'm going to do that I would do that by going and creating a calculated field. Just like you will have done in the previous section of the tutorial 4.4 point two. I've already done it here so I'm going to edit that one. And you'll see this formula is exactly what's written in the. In this tutorial, you can enter that formula ran. And then what I'm going to do is I'm going to

20:25

now calculate this table highlighted, where I've got region. Each row is a region and the columns are the years.

20:56

This time, I'm going to drag my new field into the text bar. I'm just going to. And what you can see here, is it is evolving the

21:19

sun. If I tell you that the being average. If you will now see that in the ryan two years, 2011 2012, it's always minus one. So this is confirming to us what we had in the previous. At the beginning, that those 12 missing values, doing in fact come from the 12 combinations of reason and those two years.

21:55

I'll stop there. Happy to answer some questions but I'll then I'll wait another five minutes or so before going on to the level of detail one.

29:55

Okay, let me move on to the level of detail calculation. I'm going to share my screen again.

30:19

And it may be that you already know more after this which, which gave great. This might help anybody who's just trying to finish off the level of detail. So, level of detail is quite complex. I'm not going to try to explain it all here. I'm going to point you towards just google Tableau level of detail and you'll quickly find lots of documentation and examples but I will show you a simple example what level of detail is doing is, it changes the database query, that is executed with stable changes the data that is returned by Tableau to be visualized in the worksheet that you're working with. They include or exclude versions of level of detail or dynamic expressions. So, what is returned will depend on the particular variables that are used in a visualization at any moment in time, whereas the fixed level of detail W's are the example is one where it's not dynamic it's where you the user are specifying exactly what you want. So, an expert user would use include and exclude more than they'd use fixed. But it's, it's extra complexity. Example and perhaps a little bit artificial but I've chosen it so it's still making use of this world indicators dataset, this case looking at the health expenditure in each region. And it shows how you can achieve a kind of calculation to normalize that for the region, using level of detail. So the first thing I'll do is to just not use level of detail and calculate the average health expenditure per capita per person in each region,

32:34

and remind myself what I'm doing it over the wall. It is what I want. So, you can do that by choosing region. And the average health expenditure. And it's listing the average value of health expenditure per capita. For each of the six regions.

33:09

So I've got that in one book sheets now going to show you is how you can use

33:27

level of detail to calculate the very same thing, just as an example. So this is where we're using this formula. And you can see that up phase you just need to copy this out. The first part of the formula is specify the level of detail and this is a detail by region. So that means Tableau will. When you use this measurement Tableau will return. Because think of it as an array, where there's one value for each region and the value that's returned is the average sales expenditure for all the records that are to do with that reason.

34:22

If I show you what that formula looks like. You can see it here exactly the same as it was in the tutorial document.

34:44

And the way I'm showing it. I've got the region as I had before. But this time, this pill that we're using the text is making use of the devil detail basement that we just created. And the thing to note here is that the numbers that we've got here are identical to those that we've got here. That'd be like all I've done there is prove to myself that I've done that made a mistake with my level of detail calculation

35:36

that actually what we want to do we are aiming to produce this chart here in the chart I'm just showing Middle East because that makes it manageable to read. And what we're doing here is, we are looking at the health expenditure

35:56

in each country. In the context of health expenditure within that region. So, then go to another level of detail calculation of F, which is in the first. It's calculating the average health expenditure for a given country, Bahrain Iran and so forth, and dividing it by the average for the region which came from my first level of detail calculation. I could

36:32

have actually just had one formula. And the second level of detail calculation in here. When you make use of that. Then, we end up with this as the chart. I'm filtering by the Middle East. The job is manageable. Got redone as before. We've got in the country.

37:04

And the. It doesn't matter what value you put in here I can make the average or minimum.

37:15

It doesn't change, because it's. I've already. I'm already only returning one value per region from this level of detail calculation that steps through what I'm doing today it's documented in the tutorial. It is something that is quite complex. And so you will want to spend some time working through it. And that's best done in your own time if you have queries then do post them on teams.

37:57

That'd be they'll move on to the challenges to explain where you might use those two new calculations that I added today. So these are the tableau challenges that are on the, on the data. Just to remind you where they are go to that. These challenges here. And specifically because I'm trying to get you to think about the earliest days of data analysis. Try the challenges two and four. So if I go down to challenge two. You'll see. This uses data from the car parking fines are talked about in lectures. In this challenge two. Part of the issue is, there are nine nodes. Just going to zoom in, which you can see down here in the visualizations and this is just like the situation. You found when you looking at the co2 emissions. And the question is, where did those nodes come from. If you showed me this particular visualization in a presentation. If I was your boss. And I noticed though and those are some of the day, who was flagged as being missing, and you hadn't explained that to me, I'd be asking questions and I'd expect you to know the answer. If you didn't

know the answer. I'd tell you to get out of my office, and go and find out and not waste your time until you've done your analysis properly. It's really crucial thing to do. So, to find out why there are nine nodes. You want to use. Same type of approach, as we did with the. If no calculation. And I'll let you think about that briefly challenges three and four challenge three is quite easy. It's just about visualizing data to see where there are gaps.

40:20

So you should be able to create this zoom in a bit. You should be able to create this chart,

40:26

which is showing the number of fines over time that each car park and you can see some obvious gaps here and a strange car park where there's only appears to be one data point. And then lastly, this challenge for is quite involved on it will require you to think about how you can use if no, and also make use of level of detail. Start by seeing if you can answer this. They work out this is challenge two.

41:10

And

41:15

you do get, if you do work it out then fix something in chat. If you don't work it out, I'll, I'll give you a hint just before we finish today will stop sharing temporarily. Ruben I would just a quick look at that and then I'll replied, good, good to ask these questions,

43:14

just don't know what you're getting Ruben. You say you're not getting the correct dates on the x axis with challenge three. What are, how are your dates appearing. Okay, December, January to December no yeah I just got to experiment and try and recreate what I think you're getting and then I'll show you how to fix it.

44:13

Are your pills. is your year pill blue or green. Yeah. Okay, good. So, this is why I like these live sessions because one one student encounters a problem, and I can then explain the solution to everybody and everybody else is there for finding out about something before they get the problem. I'm going to share my screen again and show you around sir. Let me start from scratch. By going to just drag the issue date into the columns. And I'm going to drag the number of files into the rows. And because they in the challenge. It was a challenge. challenge to do that. Notice that it's the granularity of months. Is that expanded this out. Sorry, hang on. I suspect that what you've done is chosen month like this.

46:03

And this comes back to is actually in the tutorial link in Section 4.4 point one about five and date. If you click on the down arrow. You'll notice that there are lots of lots of options for the dates. All these options up here these are blue ones. And all these options down here with continuous data. These are where the fill turns green but notice the difference, the month they're saying just may whereas down here thing may follow by 2015, the year. The what we've done at the moment is we're showing the number of fines and how that varies in different months, irrespective of the year. If we change it to this, we're now seeing how the number of fines varies from March 2013 across the march 2016 in months. But this time, Having is varying continuously from left to right. The key is in the color of the pill. Okay so Ben. I see your query. The question there is going to be what level of war calculations you've entered. Can you paste your. An example of your calculation. Okay, I'll, I'll try to give you a steer in a second. I'm just going to do with challenges to the set of wood and I realized we're getting close to the end of this session. In order to investigate these nine nulls.

49:16

Bring a workbook, in here. I have created a new variable, which looked like this. And in this I've got two functions embedded within each other. The first one

49:46

is presenting returning minus. 999,999. If the value in spaces is missing is not the second one double if it's a, it's a concise, if an else statement is going to return true or false, such that if spaces is positive, it will return true. And if it's negative in other words if space is invisible, it'll return false. And with a, with that, if I then just calculate the data showing the. Let me show you what how I did this. First of all, drag spaces and then I'm going to

drag. I want to save the level of detail of of the car park that I'm clicking on car park, and I'm saying the measure count the number of distinct car parks. And then I'm going to complete it I want to infect. Those the manipulations I've made these nine falses, then correspond to the nine now that I've got down here so I can go to my boss and when he asked me that question I could walk and present and tell him why there's no data.

51:54

Talking about ban on this pulling up the notes of mine. The way that I did it, then, is a bit different to the way that you did it, and I'll give you a did it in three stages.

53:00

And in that I used. The first thing I did was to find out the earliest issue date from each car park.

53:50

The first day when you have a level of detail statement. What you'll find is a maximum of issues that looks correct for the last issue date. Just wondering whether you're missing from brackets, in the, in the double if statement. Let me show you what mine

54:29

looks like. So I'm going to show you my last issue date level of detail calculation, which looks. I think is the same as the ones you have been.

55:03

And then, I now looking at the equivalent of the double if statement that I came up with. And the main thing to note is the the I've got. I'm using quote marks for my day. It also knows that I've got my date. In a year first format. I can't tell you off the top my head, what Tableau would do with what you entered in with the Try it. Try it with what I entered a first of all, in quotes, and then with the date in year first all map. You'll also notice that when you've got greater than 29th of February. I've got greater than or equal to the first of March, so those are equivalent that hopefully that will sort you out. I realized where if I'd run again, that's, that's fine that shows the value of this next week practicable we'll be throwing some data around you, and asking you to make sense of it. In that you might

want to make use of Tableau but you're also very welcome to make use of computational methods like pandas R and and so forth, you hopefully you'll find found this introduction to Tableau useful. And if you have more queries do post them on the module team channels.

56:40

Regarding assessments. I have sent my assessment data to the school. The school is gathering that for all the modules. And when the school has confirmed that they are happy with the data that I proposed for assessments. Then I'll be able to tell you when assessments take place. I cannot tell you yet. Oh Gods good news Ben, so So was it the quotes or the order of the

57:08

year first order. Got it. Yeah, okay, is what they need to know that guy stuff, obviously everybody on today's date.