[00:00:00] so our client is an e-commerce company that wants to improve the efficacy of its ad spend by targeting customers that are most likely to become high value long-term customers so currently the client spends money on ads across four different marketing channels and it wants to know which are the most effective [Music] we have two tables that we're starting with the first one the first one is the attribution table it's got three columns there's a session ID which is of type string there's marketing channel also type string and then there's purchase value Okay purchase value which is a high float[00:00:46] yeah float okay and then the second table is a user sessions table use intercessions this one also has three columns okay first one is session ID which is a type string okay there is ADD click timestamp which is of time type date time okay and then the third one is user ID which is also a string okay great all right so let's start with what is the average purchase value by marketing channel okay so I think the first thing I'm going to do is just take a look at these tables just to like lay on eyes on them[00:01:36] usually it's a good idea when you're dealing with the data you haven't seen before so I'm just going to do select star from um this attribution table let's just take a quick look here okie doke okay so we're seeing session ID marketing channel okay we have SEO paid social strategic Partnerships and paid search look like before and then purchase value here it seems like we have a lot of zeros and then um and then varying values here so what I'm thinking is going on here is the user is clicking on an ad in a particular marketing channel um and then you know most of the time they're not buying anything when they come through but sometimes they're buying stuff and when they do you know it's various sizes here okay that seems[00:02:23] to make sense if we look at let's see user ID here user sessions I run this we have user ID add click timestamp okay so pretty much it's pretty straightforward we have when the user clicks on the ad what time stamp are they clicking on it who the user is and then the session ID to connect the two tables okay so now the question is what is the average purchase value by marketing channel so yeah we have a marketing channel and purchase value both in the attributions table here so it seems like that's the only one we're going to need and since we're just looking for average purchase value that seems like a pretty straightforward group buy so we want to do a marketing channel[00:03:13] and then we want a average of the purchase value from attribution the um here we go attributions table and then we just want to group by marketing channel because that's what we're grouping by and in order to aggregate up the purchase value and then for good measure let's order by purchase value so that we can see have a better just a better database Biz there all right so what we're seeing here oh I should probably have given this a name as average purchase value there we go so what we're seeing here is SEO as a marketing channel seems to be coming out on top[00:04:04] um and paid social is just a little bit behind um hey social I'm guessing that's like Facebook Instagram ads and paid search here is you know way way down there people seem to not be buying much just when it's like on the top of Google page search so um yeah that's what that's what it seems like so far so if we're thinking about you know the overall goal of targeting uh our ad spends better it seems like SEO is probably the way to go maybe paid social as well okay this is a great start so what percentage of Link clicks convert to a purchase for each marketing channel link clicks convert to a purchase for each marketing channel Okay so[00:04:50] for this um we want to looking at conversion rates is a somewhat different question but I think we can actually keep this in the same table we just add another another column here so if we're going to go with conversion rate basically we want to know is our purchase value greater than zero or not um and if it is if we make that binary one or zero then we can take an average of that and that will give us the conversion rate so we can do a case statement here a case when purchase value greater than zero then one else zero else zero and and we're going to want to take an average of that[00:05:36] and then we'll give this a name kind of version rate and for good measure let's just do that descending as well and if I run this okay now we can see we have a conversion rate as well um and while it is in the same order it's interesting we see that SEO actually despite having a pretty small advantage in purchase value it has a 50 higher conversion rate which is somewhat interesting yeah yeah that is interesting for us to know okay so what is the customer lifetime value for each user okay lifetime value for each user[00:06:21] so if we're looking at customer lifetime value um now we're bringing in we want it for each user so we're going to need our user sessions table um because we have our user IDs here we have our purchase value here so if lifetime value I assume we're defining that as the sum of the purchase values in the whole data set for every user is that right mm-hmm um so then what we're going to need to do is join these two tables and aggregate by user ID and sum up purchase value so I'm gonna get rid of this here so if we're going to do our user ID we're going to need U dot user ID U for the user sessions table um and then I'll need a I'll need the[00:07:10] sum of a DOT purchase value from user user sessions table here as you and we're going to do a inner join here with the attributions table as a and then we're wanting to do this on a DOT session ID equals u.session ID because that's what connects the tables and here we go we then we need to group by U dot user ID and I believe that should[00:07:58] do this for us let's take a look uh oh I should probably also for a good measure order by the purchase value descending so we'll run this oh and I should probably give this a name as clv and let's get rid of this limit here also which is like automatically there okay so we're seeing that our top user here has spent like 3 500 dollars over their lifetime um and we have looks like fifteen thousand total users um in the data set uh we go to the end here we have fifteen thousand so a lot of them have clv of zero if I go to like say page 50 here or something see what happens yeah so okay we got like[00:08:51] fourteen dollars here um so it seems like there's sort of like this Peak at the top which is pretty expected you have you know your very high value users and then it just sort of like has a long tail after that yeah yeah which makes sense yeah um okay so we have our list of uh high value users now let's say we Define high value a little bit more granularly and we say that those are the users that have a clv of greater than say a hundred can we identify our highest value customers so yeah that shouldn't be too difficult so we can basically say since we're doing um clv is coming after we do our group buying here since it's aggregated we're[00:09:37] gonna need to do a having Clause here and then we can just do um a job purchase value greater than a hundred and if we run this here now we can see okay now we only have 952 users instead of the 15 000 we had before if we go all the way to the end here we can see okay everybody over a hundred yeah okay so there we go nice okay so okay moving on so first touch attribution is an analysis in which we determine um the marketing channel through which a user first encounters our product okay can you do first touch attribution for all of our only high value customers[00:10:23] attribution for all high value customers and this is um what marketing channel they first encountered the product yes okay all right so this seems like a more complicated question um let's let's break this down for a second um I think I'm going to break this into a couple of different queries here um in order to make it a little bit simpler the reason being that we need to we need to do a couple of filters so first we need to figure out uh just who all the high users are which we've already done here so that that's easy enough um and then we need to figure out what the first touch was for every user and then once we do that because the issue[00:11:14] here is that we have these user sessions and attribution table both of these have and these recessions table every user had might have many many different ad clicks and we only want to be using be focusing on the one that's first and then we only want to put on the marketing channel for that one so what I'm going to do is I'm going to make a table here and we use this with clause which allows us to basically create uh sub queries outside of an actual query which is a nice little um a nice little feature here so we can do high value as write this and just to show how this works so now I should be able to do select star from high value here and if I run this we should be getting the same thing we[00:11:59] got before okay 952. um so the next one is I'm going to want to make a first session table where I just get all the um the time stamps of the first sessions for every user and so for that um we're going to need basically just this user sessions table because we have user ID and act like timestamp here and I just want to know what the first one was for everybody so we're going to have select um user ID add click timestamp from the user sessions table and we're just going to group by user ID and then I should be able to do this again select[00:12:47] star from first session and if I run this here oops oh [Music] ah I forgot minimum you need the minim ad click timestamp here because we're doing an aggregation so we run this and that's so here we have the first and I'm going to call this first touch here so we have here every user ID and what the timestamp was for the first time they ever encountered our product the first ad they ever clicked on okay okay so now that we go ahead oh that's all I'm good so now that we have this in place we're ready to go[00:13:34] start solving this problem so um what we're going to do here is we know where we need to end up which is we want user ID and we want marketing channel as our two columns so we're going to start with select and I'm gonna I'm gonna be using the first session table here just to make things simple so F dot user ID and then a DOT marketing channel that's where we're going okay now from first session as F the first thing I'm going to want to do and I could potentially do this in a different order but um I'm going to start by filtering down so that we're only looking at our high value customers because this first session table has all our user IDs in it so we're going to look at inner joining[00:14:21] our high value table as h on F dot user ID equals H dot user ID and the reason that this is going to work is because we're doing an inner join right it's only going to let the rows survive where there is a match between user ID in the high value table and the user ID in the first session table which is only going to be where we have high value users so this should we did this right I'm just going to comment this out real quick we did this right we should have 952 rows and otherwise it should look just like there we go 952 rows okay so that's good now coming back here The Next Step here is we have to get to marketing channel all right this is where we're trying to go[00:15:09] and the issue is in order to pull in this attributions table right now we don't have anything that connects right what we need is session ID in order to bring in this attribution table and so actually we're gonna have to bring in the user sessions table first so that we can get session ID and we're going to have to bring in the session ID that matches the user ID and the add click timestamp because we have every user ID a user ID might have multiple uh time stamps right and uh timestamp might have multiple users and so we need both the session ID that connects to both so we're gonna have to do here it's another inner join of um our user sessions table[00:15:54] on U dot user ID equals F dot user ID and U dot add click timestamp equals F DOT first touch okay and then lastly we're able to bring in our marketing our marketing channel with the attribution table um on U dot session ID equals a DOT session ID so if I run this now hopefully no errors there we go so now we have every user with a marketing channel that they first encountered our product and we have it for this fantastic 52 users who are the the 100.[00:16:40] awesome okay so now the last last step for us is can you now calculate what percentage of high value customers came from which marketing channel originally I value customers came from which marketing channel originally okay so right now we have the marketing channel for all the high value customers so this is just going to be an issue of aggregation here um and we're going to want to aggregate by by marketing channel so I'm going to switch this over here real quick put a marketing channel first um now the question is we what we want is they're asking for a percentage here so um we're going to want basically a count[00:17:27] divided by the total so I'm going to start with count of f dot user ID and then we're going to divide and we know it's 952 but for the sake of not hard coding um I'm just going to take the count of everything in the high value table because that's the same that's all of our customers here so I'm just going to do a sub query here select account star uh from I value and then lastly because this is um we're doing division here SQL doesn't like it when you divide an integer you need to divide a float so I'm going to cast this as a float first okay and then we need to do our group by so group[00:18:14] by a.m marketing channel and as always for a good measure let's do come on an order by um just take this or actually I could just do this pretty simple descending and if we run this oh I forgot to give it a name uh percent uh high value we can see now that actually paid social is responsible for 42 of the high value customers that we're getting here um which is really interesting so before we were seeing that SEO is on top and now we're seeing it social actually as far as getting the customers we most care about who are going to be providing[00:19:13] us the most value in the long term this might be the place that we want to focus our ad spend paiser is still at the bottom here and SEO isn't terrible but um significantly lower than uh than paid social yeah yeah I'm sure Marketing Executives would love this insight absolutely thank you so much ben um I think we can conclude over here I'd love to hear your thoughts on this like what are what are some things that you think you did that you would recommend other people do as well in an interview that they encounter like this one yeah so I mean I think it's one thing that's really good is to take good notes when they are giving you what the question is make sure you're not missing anything asking clarifying questions um so I believe that I asked a question[00:19:57] about I forget what it was actually in this moment but I think one of the early questions early questions I asked uh um oh I think it was clv that we were just talking about the total percentage the total sum within this data set has the definition um making sure you understand the tables and the types here and then as you're going through it's good to State what your intention is what your plan is for your code before you write it how you're going to go about solving it and then as you go through um saying what you're doing and then when you're done saying what you did that's the general presentation guidelines um yeah uh and trying to make it structured and making sure that you're bringing it back to the business case at any point that's[00:20:44] possible you always want to especially as a data scientist there are so many data scientists and analysts who are just focused on the numbers and and forget what the business case is and that really is something that companies are looking for and sort of help set you apart so you want to make sure that you're always keeping in mind who the customer is who the client is why we're why we're doing this and how this is going to help them or not and so when you get to your results at the end you can say how this is going to be implemented and how it might be used absolutely yeah I definitely noticed some of those things that you were doing as well and they really helped you um stand out I think in this interview there was you were definitely bringing me along for the journey with you so I understood every step of the way okay this is what Ben is thinking this is what he's planning he's not done[00:21:29] answering the question this is just like part one he's setting up for what he needs the ingredients that he needs to be able to answer the question um the other thing I noticed was you definitely did reference the the question that I had asked a couple of times you were like okay but this is you know this is probably what you're looking for this is like the underlying question that you're trying to answer based on um based on what the context is at the interview and of the question that I originally asked and uh and then I like that you called out a couple of things like just best practices in your code as well you were kind of calling them out as you went along like don't hard code values try and and why why is that important these are like obvious little things but because they're so basic like[00:22:14] sometimes they get left out um and it's always reassuring to the interviewer to know that okay I'm dealing with someone who knows what they're doing um and they they understand these best practices and use them on a pretty regular basis as well yeah um so yeah thank you so much this was super helpful and I'm sure this is going to be very helpful for everyone who's watching at home yeah awesome great thank you very much and this is a fun exercise and uh excited to see how this helps people absolutely and best of luck with everyone who's preparing for their interviews and for you as well if you're ever preparing for those in the future best of luck all right thanks everyone take care thanks so much for watching don't forget to hit the like And subscribe buttons[00:23:00] below to let us know that this video is valuable for you and of course check out hundreds more videos just like this at try exponent.com thanks for watching and good luck on your upcoming interview foreign [Music]