we have to find the unique work base for the employees on which they have worked so it says that we have a table containing information about the time that employees have worked on a project there will be employee ID the project they have worked on and the start date and end date of each project we need to write a query to find how many unique calendar days each employee has worked and we need to order our query by the employee ID hello everyone we are back again with another mock interview at interview query my name is Ravi I am a staff engineer at mang company I have previously worked for other mang companies also and I specialize in different techs including data engineering including backend engineering and software engineering as usual today we have samri with us for this mock session samriti would you like to introduce yourself thank you so much Ravi hello everyone my name is samri wats a little background about myself starting with my undergrad I am an engineering by my undergrad degree I completed chemical engineering back in India in 2021 post that I was introduced to the field of analytics and big data when I joined a Consulting Pharma consulting firm Zs Associates back in India I worked as a decision analytics associate for 2 years from 2021 to 2023 August 2023 in Zs where I was primarily interacting with the top pharmaceutical Fortune 500 pharmaceutical client and helping them with their forecasting related operations and business queries related to product launch lach Market entry and building dashboards for brand performance and brand planning post my stint at Zs I decided to pursue analytics and data science uh further and I joined P University as a master students in business analytics and information management program in August 2023 and So currently I'm a grad student at Purdue and I am set to graduate in August 2024 I'm looking to find roles as data scientist and business analyst in different companies across us awesome thank you for that s so let's head over to the question today we have a very good question which has been asked in several top tier companies you have the examples in front of you as you can see the table example is there that one employee can work on multiple projects and the start date and end date can overlap as you see and in this example the employee who has worked on Project one and project two you can see the start date and end date there is an overlap from 3 to six between both the projects so the unique work days comes out to be seven now that's the question you can get started the stage is all yours okay so my first idea when I read this question was to count the number of days for each employees like take the difference between end date and start date and count the number of days they have worked but then as the question mentions that there could be an overlap so for each employee ID if you just Group by employee ID and just take the number of days as a difference between end dat and start date dat that would not be accurate as it would count repeated days when a single employee was employed in a different project so I guess we I can try writing the query so starting with the basics we have to Output employee ID so we can write that employee ID and then we'll have to Output the number of days work so that would be a numerical value and that would be a count so we can just have count keep this blank for now because we can decide which column to use so this could be as this we are taking from we'll create a new column in the projects table and we will Buy employee ID and we'll order it by employee ID as the question says now we'll have to find out how to get the unique work days so so I think we can we can have all the dates list Ed out for all the projects and employee worked for example for employee ID 1 and for project ID 1 it worked from 1st January 2024 to 6th of January 2024 so we can have like 1 2 3 4 all the states listed and same for the project ID 2 so here we'll have all the dates listed and then we can take the unique or the distinct count of the employee ID and the date pairs so that will basically generate us unique combination of dates for each employee that it worked on and later on we can take account of those unique number of days for each employee ID so this is my idea so if I get your idea correctly you want to First list each of the dates for each project actually and then maybe remove the duplicates from them that's what you are going towards I believe so I won't like here I won't use any Group by or order by project I would just list down all the dates like starting from the start date I'll just list on all the dates between start date and end date every date and we can create a temporary table of all those dates and then we can have another table which has the unique combinations of employee ID and the dates so that will give us like unique pairs of every employee and the date on which it worked so suppose like employee ID 1 and 1 January employee ID 2 and 2nd January employee ID 3 and like 6th January so this will count the factor of overlap that was happening before because it will generate the distinct dates so one thing I'm confused here is that when you use employee ID and the dates so for example in this in this example that is given here employee ID has worked on two projects so if you take employee ad1 I mean you would get two entries for 3rd of January 2024 for employee ad1 I mean how is that unique is what I'm thinking so so what I'm thinking is we can have a table just for the dates first which is not having the employee ID okay where the date starts from the first start date and it has an interval of every day till the end date and we can just have the dates in that like the dates column and from there we can like write another query to fetch the employee ID and their corresponding work date okay okay so that will create like we can have one table for dates and we can create another table which is fetching the employee ID and their work dates so there it is creating unique pairs not unique pairs there it's creating basically Pairs and then we can take the count of it okay okay and uh if you find two similar pairs like what will be your decision so for example as I said that employee D1 you would get two pairs of for 3rd of January 2024 right because employee id1 has worked on 3rd January 2024 for project id1 and 3rd January 2024 they have worked for project ID 2 also so if you do like that you would get two pairs for third January we will take the distinct thing yeah okay okay I think you might be heading in the right direction I would love to explore more on this but yeah you can move move forward yeah okay so we can then we can generate I believe you are are writing this with respect to keeping my in in perspective which quy are you no I I was writing with post J yeah fine fair enough okay so here I guess are you able to see my question yes I'm able to see it so basically with the first sub with the first CT expression range of dates that is listing out the individual dates for each project like we use the series function like it is creating a row for each date like an employee worked on a project and the second CT expression is basically taking the distinct pair of for the employee IDs and the date worked in order to overcome the problem of an employee working on a two different projects on same date so that will help remove the duplicates in in that situation and the final statement is basically listing out all the employee IDs and it is counting the date of like the counting the dates when with the employee worked to generate the number of unique days and then we are grouping it by employee ID and ordering it by employee ID do you mind um explaining a bit more on the series function that you're using like how it works and how it is working in this scenario actually how is it helping in this scenario yeah sure so here basically my idea was initially to have have all the dates listed right like initially as we thought we since there is an overlap between date start date and end date for a single employee because it has worked on multiple projects we'll have to list out all the dates like all the dates which are there in this database for all the projects that an employee has worked on so generate series function is basically generating all the dates between start date and end date so for example if I take the given example of the table if the start date is 1st January 2024 this series is generating all the dates in between from the start date till end date including the dates of the boundaries like the start date and the end date and it just takes interval one day is specified so that it takes into account each day and there is no it's not skipping any dates as such in between so yeah so it takes both the start date and and it like it's inclusive of them right yes okay fine so this will generate you all the dates on which the employee has worked for IR respective of the project it will generate for the employees right yes yes yeah that's my idea and then since we have all the dates listed now we can like take the unique Bears of employees and the dates on which they work so the second query is doing that work for us that it is taking the distinct pairs of employees in the dates that they worked and that will deal with the overlap issue that's happening okay okay and finally how are you counting the unique T so in the final select statement it is selecting the employee ID and it is counting the date work column which was created previously so now since we have for the second Sub C that we have written we have a table which lists out for each employee what was the date it worked like for one maybe it was 1 January 2 January 3rd January so this final select statement is basically counting like seven days five days the number of unique days that it worked each employee so it is basically counting those dates and here we have just used count and we have not used distinct count because we already generated the distinct dates in the previous query fair enough okay all right okay yeah would you would you like to try this out yeah sure syntax second is not going on okay looks like it is passing the initial test cases fair enough yeah would you like to see it against further test cases maybe just try to submit your solution if it gets accepted okay okay cool it gets accepted so one of the things I would like to understand while this is fine I mean I understand this is working but can you tell me you know a bit or elaborate a bit more on the idea that is there is there any other approach which is striking your mind or like how did you come up with this approach that just first get all the dates and then find the unique ones in that like how did you arrive at this approach I'm trying to understand the intuition of this approach here so like as I mentioned during the start of the interview the first idea when I read the question that came to my mind was that we'll have to without thinking about subqueries or CTS that we'll have to get the number of days each employee worked so that that can come with the difference between end dat and start dat but then that would generate the the duplicates because of the multiple projects a single employee had worked so my second approach was like my second thought was that we have to basically get all the unique days as in there shouldn't be any duplicates in the dates for each employee irrespective of the projects it worked like all the works that an employee has done whether it was in Project one or project two we have to get basically all the dates unique dates it doesn't matter that on a single day an employee was working on five projects or one project I we need that dat one one time because the employee was enrolled right so from there it started that we have to just get like list of all the dates not just that like the difference or not not like picking for each project wise we just have to get the dates corresponding to each employee that it worked on so it could have also been the like we we could have set the start date as the minimum date and uh like and like end date as the maximum date and we could have taken all the like the gaps between the Max and the Min of that we could have also done that like I did not try that approach but yeah that also came to my mind that start date is the minimum date the first date that employees started working on and we can have a set a range of Min and Max and then calculate all the dates that come in between so that will give us all the unique dates that is coming in between and irrespective of the projects it worked on okay cool fair enough I think yes I get the idea just wanted to dig a a bit more into into the details of the intuition uh and uh why you took certain decisions so I think that was well explained and given that this query works it obviously is a working Solution that's that's what I will accepted this as uh now since we have got to the answer for this and it was a pretty I would say it was not so straightforward also but Kos to you that you thought of the areas of the duplication will occur and how to remove that so there are certain observations that I have made out of this I would say solution and maybe we can discuss a bit on the feedback if if you're okay with that yeah definitely sure so one of the aspects first aspect of navigating through the problem so I think you read through the problem and you understood and while the examples and everything was pretty clear for the problem I would still suggest or recommend that from an actual interview perspective maybe still try to ask a few questions to the interviewer just to clarify the question a bit more I understand there might be questions which are very straightforward you know the examples are super crystal clear and there is less scope to ask question but that is exactly where the interviewer will expect that given a real world problem does the candidate directly jump to the solution or do they have a mentality or do do they have a work ethics of you know discussing or clarifying it a bit before moving to the approach or deciding on the approach okay so that's one feedback there that you know it's always good to clarify or ask a bit of questions even if you have less scope for it but just ask a bit few maybe couple of questions just to give a good impression to the interviewer that's one thing and I I I should not say just to give a good impression but many of times actually asking questions help you reestablish or reconfirm that you are understanding the question exactly how the interviewer is understanding it so that's the intention there I should say now moving into the solution I think you adopted a good approach of explaining the solution I should say where according to the criteria of the question you first started with a I would say naive approach where you mentioned that ideally we could have done something like the difference of the dates the start date and end date and then you recognize you highlighted what would be the problem with that and essentially then your further approach was focused on eliminating that problem problem now this is a really good way of describing the solution to the interviewer I I will give you full marks for that where you started with a very raw solution very raw thought process and then highlighted or identified the problem with that raw thought process and then worked towards optimizing or getting to the better solution from there and I think this is a really good approach a a good example I should say whoever is learning from from this mock session that this is how you should be doing it even if you are very aligned or inclined towards thinking for the most optimal solution for the best solution but the interviewer more than the solution or just seeing the code running or the query run I should say they are more interested in your thought process like how did you arrive at that solution and I think you did a good job of explaining that so good job done there now while we decided on the approach and I think it was very much you know your approach looked fine to me then we moved ahead with the solution this is how usually is a good way of doing this in the actual interview also that you first discuss the approach with the interviewer and then move ahead with writing the query now uh one suggestion there one another feedback that I noticed there that your communication is also something which is noticed by the interviewer and in the initial part of the interview while everything was on point you were thinking out loud and and very communicative of your of your ideas I think to the part when you started writing the query uh there is hope that we can still look to be communicative as we were earlier and how do we do that that while I'm writing that query I keep on explaining what I'm doing in this query what why what this query does in particular okay yeah so the more you are communicative the more you make it easier for the interviewer the better you you will score in the interview that's as simple as that and obviously for the intention that you want to avoid those awkward silences because silences might sometimes you know be perceived differently by different interviewer so just for that sake also so that's just one another feedback that we can look into there and again your query worked you were able to identify the the exception also and correct that which was also good fine enough another suggestion which I will have is when you're using some Advanced functionalities of SQL like CT in this case for example like the CDs function that you used so take this opportunity to plaunt your knowledge a bit the fact that you are using them means you are knowledgeable about them and this should be your opportunity to highlight or explain about those functionalities that how do they work and how they are assisting you in your query actually okay okay and proactively doing that is going to create a a very good impression on the interviewer if you proactively do that so yeah all in all those are the feedback that I have and yeah that's it that's that's all the feedback that I have any questions that you have some with thank you so much for the feedback it's it's really helpful because I have an upcoming interview as well so I guess I would be able to leverage all these feedbacks in that interview and improve my interviewing skills I don't have any as such questions as of now because of the related to the SQL problem that we did but in general like since I will be approaching I will be mainly interviewing for the data science roles so in generally have any specific feedbacks for data science related interviews where they do not just have the technical rounds but also like sometimes business case based rounds so if you have any idea or any suggestions on on feedback on those it would be great to hear on that so probably how many rounds do you have for business use cases like if you have some insights on that I mean are there are there are there more than one rounds for business use cases or are you just like a single round for a business case based interview see one thing which I can suggest for sure for data related interviews is that data science deals or is more I would say more looking forward for the domain specific knowledge also more than the data science technical part of the things it is also more aligned more focused on how good you are with different domains and that's where this business use cases kind of questions comes into perspective they are looking to I would say looking to not only test your technical acument there but also your business Acumen on on identifying or on making business decisions and your communication abilities these are the primary things that they are looking to judge you on okay whenever a business use case comes first of all like from the preparation perspective you can look at some popular business domains from the preparation perspective and just try to take an idea just try to understand how different domains work what are some specifics with respect to each domain for example Healthcare domain Aviation domain e-commerce domain things like that just get an idea about how different domains work essentially and then when you are presented with the question I would I would also suggest that maybe you can understand a bit about primarily like try to break down the problem like essentially as simple as what are the inputs what are the outputs what are the goal which is expected and all okay and all the thing which will happen in between so kind of try to take it from a very high level and then try to dig deeper into the things is what my suggestion would be okay and then the clarifying question are non-negotiable because the more you try to understand the more you try to kind of get a Clarity around what are the specifics around probably starting with what the system is expecting like maybe what are the inputs which are expected when we are trying to solve this so maybe it's it's a customer data so where is the data coming from like what is the what are the sources of data what are the different types of data that we are dealing with okay trying to understand or ask these kind of question clarifying questions and then similar questions for the output that what is expected on the output what are we trying to do with output or what are we going to use it for things like that probably so trying to establish the scope trying to understand clarifying questions around these parts that is how you break down the problem from the very top level to the simpler one module kind of thing at a time like divide divide the problem kind of thing and then try to address each of the parts individually and then obviously you you go on from there so I mean U technical wise I mean I will not go into much details here because that's definitely I would say that's something which you which changes from every question to question or I would say use case per use case but in general if you want to follow kind of a template this is something which you can follow where first have a generic idea about different business domains okay and then try to break down the problem and even if for some business use case you might not have an idea that I am not much familiar with this business domain or this kind of use case then this is how you exactly approach it that you try to break it down what are the what is the input what is the expected output like what is the goal we trying to achieve and then try to ask some more clarifying questions around both those areas just so that you understand what you are trying to solve on that given day okay that will be a template that I will suggest you to follow all right that completely makes sense and thank you so much it's it is definitely very helpful and I will try to implement it in my preparation and the interview as well absolutely my pleasure and that will be all from my side thank you so much for checking out time and for folks who are interested in solving such interesting questions and preparing for your upcoming interviews at Big Tech and all the different companies which are interested in data related positions check out different questions and the interview query platform it's there to help you out that's all from my side thank you