



# IIT Madras

## ONLINE DEGREE

**Computational Thinking**  
**Professor Madhavan Mukund**  
**Department of Computer Science**  
**Chennai Mathematical Institute**  
**Professor G Venkatesh**  
**Department of Electrical Engineering**  
**Indian Institute of Technology Madras**  
**Iterations using combination of filtering conditions**

Professor Madhavan Mukund: So last time we looked at this question of filtering, so we said we would go through the cards and based on some characteristics of the cards, some property of the card, we will decide whether it is useful or not. For instance, we looked at pronouns, we looked at verbs we also said we could keep track of both pronouns and verbs at the same time because something is either a pronoun or a verb. So, now let us do something more interesting.

(Refer Slide Time: 0:41)

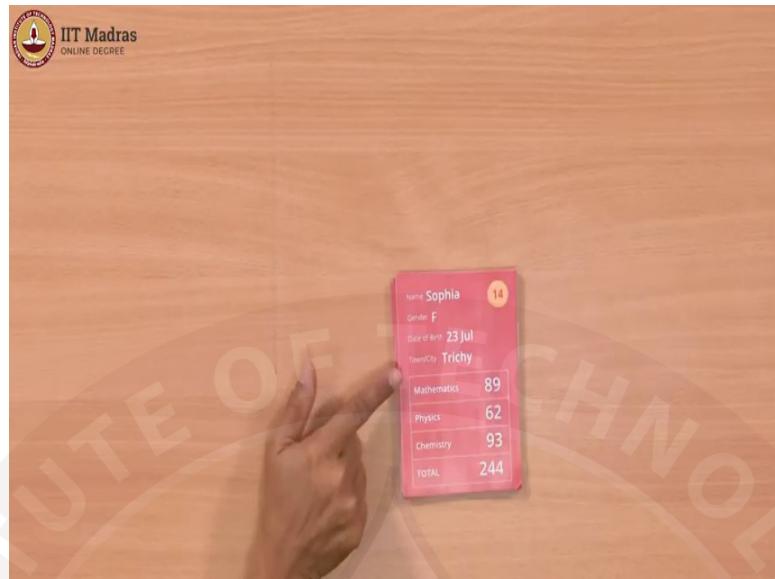


Professor G. Venkatesh: So I am going back to this data set, classroom data set. Now let us say in this data set I want to find out all the girls, who not just want do I want to find out how many girls are there, but I want to find out how many girls are there from a specific city, let us say Chennai.

Professor Madhavan Mukund: Okay.

Professor G. Venkatesh: So how many girls are from Chennai, how many Chennai girls are in this...

(Refer Slide Time: 1:01)



Professor Madhavan Mukund: So there are two, so there is a gender and there is a town/city, so there two items and what we want is this should be female and the town/city should be Chennai.

Professor G. Venkatesh: Yeah, how would we do that, I mean is there simple way of doing it...

Professor Madhavan Mukund: So what we could do first is we could first separate out all the girls, so that would be one step of filtering.

Professor G. Venkatesh: Want to try that?

Professor Madhavan Mukund: Yeah. So...

(Refer Slide Time: 1:24)



Professor G. Venkatesh: So here is a girl, so I guess I put it in a different pile, here is also girl, girl, girl, boy, boy, girl, girl.

Professor Madhavan Mukund: Girl.

Professor G. Venkatesh: Girl, boy.

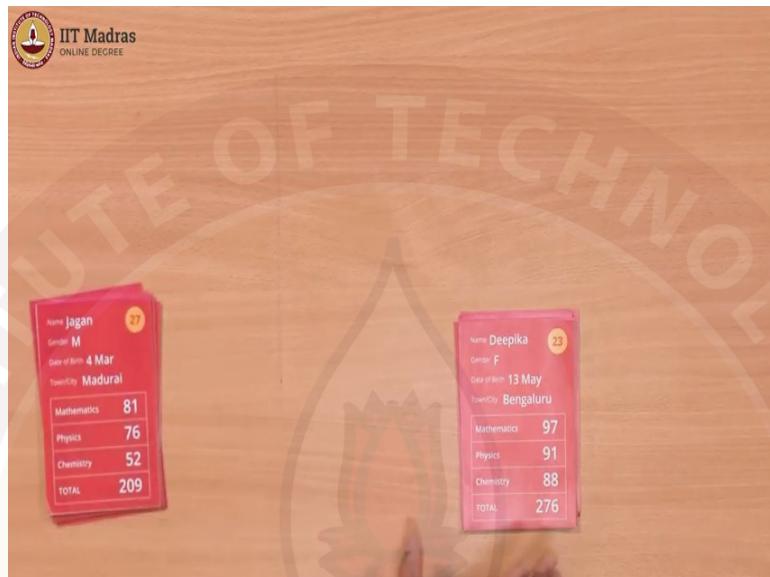
Professor Madhavan Mukund: Boy, boy, boy, boy, boy, one more girl, boy, boy, girl, boy, boy, boy, boy, girl, boy, boy, girl, girl, girl. So now we have filtered this data into two piles, everything on this pile is a boy...

Professor G. Venkatesh: Not useful...

Professor Madhavan Mukund: Everything there is the girl...

Professor G. Venkatesh: This is one which we are interested in...

(Refer Slide Time: 2:07)



Professor Madhavan Mukund: Now, among the girls, we want those who are from Chennai. So we move this aside keep it separately and now we want to go through this again look at the town/city thing.

Professor G. Venkatesh: Yeah so...

Professor Madhavan Mukund: So this is not Chennai, Bengaluru, not Chennai, this is Chennai, this is Chennai, this is Chennai, this is Chennai, not Chennai, this Chennai, not Chennai, not Chennai, not Chennai, not Chennai and not Chennai. So now we have pulled out those cards which are a combination of two different interesting things for us, that they are both girls and they are from Chennai.

Professor G. Venkatesh: And we can count this I guess. We shall 2, 3, 4, 5...

Professor Madhavan Mukund: So among this entire remember we add 30 students we had counted at the beginning, so among those we have found that there are 5 girls from Chennai. So in this what we have done is we have done filtering in two stages, so first we have applied a gender filter to pull out all the girls and then we have applied the city filter to pull out all the Chennai, I guess we could have done it the other way, we could have first pulled out all the Chennai people and look for the girls in that.

Professor G. Venkatesh: Should give the same answer, right? Because the same cards would come out.

Professor Madhavan Mukund: Yes...

Professor G. Venkatesh: First, we should have filtered the Chennai.

Professor Madhavan Mukund: Yeah, but supposing we did it in a one step, can we do it in a one single, so what I said is instead of going through and first explicitly pulling out all the girls, and then going through all the girls and then pulling out all the Chennai people, do you think we can just do it in one shot like we have done for some of the other iterations overlap this thing that we could check?

Professor G. Venkatesh: We should I guess that would be I mean just go through the cards in one iteration and at each stage you look at the card and see whether it is a female, that is not enough.

Professor Madhavan Mukund: And we also need to check...

Professor G. Venkatesh: So we have to do, so you have to check two conditions. So you have to check...

Professor Madhavan Mukund: So both of them have to be true

Professor G. Venkatesh: Both of them have to be true, so it is an ‘and’, and of two conditions, and of two conditions, the first condition is the gender F and the second condition is this the town/city Chennai. So if it is an and of, if both of them are true, then only we select the card for counting we keep count I guess, so we have a count variable and so maybe we start with a count variable to 0.

Professor Madhavan Mukund: Yeah. So we could do that, maybe I could do that by count...

Professor G. Venkatesh: And then we go through this...

(Refer Slide Time: 4:23)



Professor Madhavan Mukund: So I will say, Chennai...

Professor G. Venkatesh: Girl, Chennai girls equal to 0. So here is a card which is a male, so it does not satisfy the condition.

Professor Madhavan Mukund: So if it is a male we do not even have to look at the city.

Professor G. Venkatesh: Do not have to look at the city, so that is interesting. So, again here it is a male, so do not have to look, male, male, so female now we look at the city, Trichy...

Professor Madhavan Mukund: So it is not Chennai.

Professor G. Venkatesh: So again, female Teni, female Bangalore, female Madurai, female Chennai.

Professor Madhavan Mukund: So now I have to increment my count.

Professor G. Venkatesh: Increment the count, another female Chennai.

Professor Madhavan Mukund: Increment again.

Professor G. Venkatesh: One more female Chennai, they all come together I guess because you put the cards together, another female Chennai and then one more female Chennai, and there is a female Madurai, Erode, female Nagercoil, Bangalore, male, male, male, male so the rest I think are not cards that are useful to us, so just going through all of them.

Professor Madhavan Mukund: So now basically in one...

Professor G. Venkatesh: So we got the count, in one iteration we kept track of a variable called Chennai girls and we kept incrementing that variable provided that both the conditions were satisfied, it is ‘and’ of two conditions, gender being female and the city being Chennai.

Professor Madhavan Mukund: So this is a very useful thing now, so we know that we can take the entire stack of data or sequence of data or cards whatever we want and we can pull out those which are interesting based on some property, in this case it was the city, for example if it was the shopping bills it could be based on whether the total is bigger than something and they have bought at certain type of item, in the case of the words for instance it could have been if it is a noun and it is more than five letters long, so we can take various combinations of conditions and filter on them simultaneously.

Professor G. Venkatesh: Is there something bit more interesting we can do, for example, can we find all those people who are born between one date and other date?

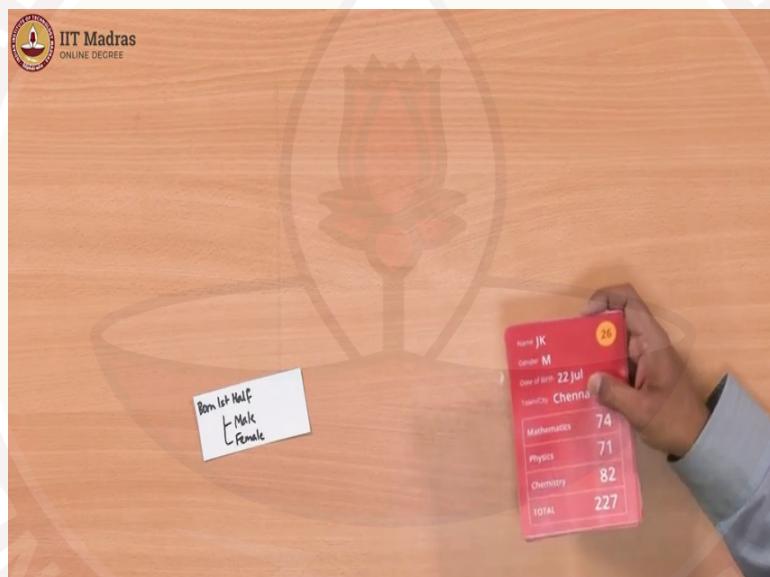
Professor Madhavan Mukund: Yeah, let us try that, so let us see if we can find all people who are born in the first half of the year, say between 1st January and 30th June.

Professor G. Venkatesh: It should be half, we do not know.

Professor Madhavan Mukund: We do not know, but let us see, is it true that, let us see.

Professor G. Venkatesh: It should be half, I mean if it is random it would be half, but this may not be random...

(Refer Slide Time: 6:46)



Professor Madhavan Mukund: So let me keep track of a count for this, so this is...

Professor G. Venkatesh: So here we are actually now not checking against two of these fields, we are checking one field only, we are checking the date but now we are comparing the date, we have two...

Professor Madhavan Mukund: So we could still do two, so we could say for male and for female, so we can say they should be born in the first half and they are male born in the, so how many boys were born in the first half of the year, how many girls...

Professor G. Venkatesh: Girls are born in the first half. So to check whether it is the first half the date has to be, so if it is 22 July is it first half or it is not first half?

Professor Madhavan Mukund: No.

Professor G. Venkatesh: It is not first half.

Professor Madhavan Mukund: So anything should be up to June 30th.

Professor G. Venkatesh: So which means that it should be...

Professor Madhavan Mukund: So we should first check the date.

Professor G. Venkatesh: It should be more than 1st January or equal to, more than or equal to 1st January and less than or equal to 30th of June.

Professor Madhavan Mukund: That is right.

Professor G. Venkatesh: So that is what it means. So let us check this this is, so this is 22 July, so does not meet the criteria. 4th March...

Professor Madhavan Mukund: Yes, and it is...

Professor G. Venkatesh: It meets the criteria because it is greater than its between 1st of January and 30th of June, so it is in that space...

Professor Madhavan Mukund: And it is a male.

Professor G. Venkatesh: And it is a male.

Professor Madhavan Mukund: So I will increment the male count by 1.

Professor G. Venkatesh: The Next one is 17th of September, so again in the second half. 30th of August...

Professor Madhavan Mukund: 2nd half.

Professor G. Venkatesh: 2nd half because it is greater than...

Professor Madhavan Mukund: After June.

Professor G. Venkatesh: After 30th of June. 6th May this one lies between 1st of January and 30th of June.

Professor Madhavan Mukund: And it is a male.

Professor G. Venkatesh: We should count it, it is a male, so it should go to the male one. 13th October again is in the second half. 3rd June is before 30th of June so should be in the first half male. 4th Jan, again male, looks like there are more males. 14 December...

Professor Madhavan Mukund: No.

Professor G. Venkatesh: So no, 30th December no, 7th November no, 30th April male.

Professor Madhavan Mukund: So that is now 5.

Professor G. Venkatesh: 26 December male, 13th May, so first time me got a female, 13th May first half. 17th July...

Professor Madhavan Mukund: No, too late.

Professor G. Venkatesh: Again no, 9th October...

Professor Madhavan Mukund: No.

Professor G. Venkatesh: 10th September.

Professor Madhavan Mukund: No.

Professor G. Venkatesh: 12th Jan.

Professor Madhavan Mukund: Yes.

Professor G. Venkatesh: Female, 16th May again, female.

Professor Madhavan Mukund: Yes.

Professor G. Venkatesh: 8th February female, 14th Jan female, 5th May female, wow, 17th November, so this does not count, 15th March...

Professor Madhavan Mukund: So I am running out of space I am moving to the left, so I will say 7 now.

Professor G. Venkatesh: 22nd September.

Professor Madhavan Mukund: No.

Professor G. Venkatesh: 23rd July.

Professor Madhavan Mukund: No.

Professor G. Venkatesh: 23rd March.

Professor Madhavan Mukund: Yes, so male now goes to 6.

Professor G. Venkatesh: 15th March, again male.

Professor Madhavan Mukund: Male goes 7.

Professor G. Venkatesh: 28th February male.

Professor Madhavan Mukund: Okay, 8.

Professor G. Venkatesh: 6th December does not count. So what do we get?

Professor Madhavan Mukund: So of the 30 students, 8 males were born in the first half, 7 females. So, roughly half the students, actually exactly half...

Professor G. Venkatesh: 15...

Professor Madhavan Mukund: 15 of them were born in the first half...

Professor G. Venkatesh: This is not bad.

Professor Madhavan Mukund: Which is surprising and off that since we have an odd number of 15 is roughly equal, 8 males and 7 females. So this is a very equally...

Professor G. Venkatesh: Looks like a balanced, looks like a very balanced data.

Professor Madhavan Mukund: So this is interesting, so now so basically so you can filter on multiple things at the same time and then when you are filtering on multiple things you can also keep track of them separately. So we have two conditions, when they were born and whether they are male or female and based on the combination we have one count for males born in the first half, one count for females born in first half.

Professor G. Venkatesh: And we have used the same pattern, we have used the iterator pattern, initial values of so we kept two variables, male female born first half variable count, that is a count so we have two variables that we are keeping track and going through the iteration each time filtering on the gender as well as the date of birth and checking whether the date of birth field falls between two values, 1st of January and 30th of June, inclusive and if it falls then we add to the male count or the female count, that is what we have done. So it is just an iteration pattern with a filter added.

Professor Madhavan Mukund: So filtered iteration, very good. So let us do something more interesting with this set, so here we have the maths marks highlighted and earlier we computed the average marks for the whole class but supposing I wanted to find out as a teacher who is better in general in the class, are the girls doing better in maths or the boys are doing better in maths. So what would be a good way to look at it, should we look at the highest marks maybe the which...

Professor G. Venkatesh: I am not very sure about that, highest marks because there could be one exceptional boy let us say, but you know the girls could general be doing better than the boys, but only one exceptional boy may be there and just by looking at the maximum you will biasing it by looking at this exceptional boy candidate. So I do not think we should look at the maximum.

Professor Madhavan Mukund: So then, if you want to look at the general trend I guess average is as good as...

Professor G. Venkatesh: Average is, average should work I think, if you find the average of the boys and the average of the girls and compare it, if the average of the girls is higher, which I think is likely than the boys, then the girls are doing better, this is what I would think is right way of doing that.

Professor Madhavan Mukund: So, we saw how to compute the average, we have to find out how many cards there are and then we have to add up the marks across those cards and divide by the number of cards but now we have to separate the cards, we have to filter the cards into those for the boys and those for the girls.

Professor G. Venkatesh: So we are doing total, we are trying to find the total average total marks for boys.

Professor Madhavan Mukund: For maths total.

Professor G. Venkatesh: Maths total, we are checking which, whether girls are doing better than boys in maths?

Professor Madhavan Mukund: In maths, yes.

Professor G. Venkatesh: So we can find the average maths marks for girls, average maths marks for boys and compare the two, that is what we want to do. So simple way would be to simply separate the boys from the girls, and then find the average for each set like we did last time, last time we did the average for the whole set...

Professor Madhavan Mukund: So we filter it out into two sets boys and girls...

Professor G. Venkatesh: Boys and girls...

Professor Madhavan Mukund: Then we process the boys separately...

Professor G. Venkatesh: Boys separately find the average, girls separately find the average and compare the average.

Professor Madhavan Mukund: So why do we need to separate them, why cannot we just do it as we are going along, so surely like we said we did it earlier, we can keep track of all these things in one single iteration, so we could keep track of the...

Professor G. Venkatesh: We need to keep track of to find average of the boys, what we need we need the total mathematics marks of only the boys and we need to find the total number of boys and then if you divide the total mathematics marks of the boys by the total number of boys that gives you the average for the boys. Similarly, for the girls you have find the total mathematics marks of the girls and find the total number of girls and then divide the total mathematics marks of girls by the total number of girls that will give the average for the girls.

Professor Madhavan Mukund: So we have to keep track...

Professor G. Venkatesh: So we see it looks like we keep track of 4 things.

(Refer Slide Time: 13:49)



Professor Madhavan Mukund: So let me just write it down, so we need say let me just break up this space into four things, so we need say the boy count...

Professor G. Venkatesh: We need the boy count.

Professor Madhavan Mukund: And of course we need the girl count. So this will tell us how many of each there are and then we will keep track of the boys sum and the girls sum.

Professor G. Venkatesh: And we are going to do all of it in one iteration, as we go through the cards...

Professor Madhavan Mukund: As we go through in one iteration.

Professor G. Venkatesh: So here there is some filtering going on because we are looking at whether it is a boy or a girl and then doing the so these few things...

Professor Madhavan Mukund: And then this, and then use accumulation because...

Professor G. Venkatesh: Accumulation because we are adding the mathematics marks and then we are counting, so we are counting, accumulating, filtering all together in one shot, so let us try it. So here is the first card, this is a boy...

Professor Madhavan Mukund: So boy is now count is 1.

Professor G. Venkatesh: And mathematics is 72.

Professor Madhavan Mukund: So the sum is now 72.

Professor G. Venkatesh: Second one is a girl.

Professor Madhavan Mukund: Girl count is 1.

Professor G. Venkatesh: Girl count is 1 and 74 is the maths marks. Third one is again a boy, count...

Professor Madhavan Mukund: Count is 2.

Professor G. Venkatesh: Maths is 81.

Professor Madhavan Mukund: So 72 plus 81 is 153.

Professor G. Venkatesh: The next card is a male again boy...

Professor Madhavan Mukund: 3.

Professor G. Venkatesh: 74 is the maths marks.

Professor Madhavan Mukund: So that is 227.

Professor G. Venkatesh: The next is girl and 62 is the maths marks.

Professor Madhavan Mukund: So it is 136.

Professor G. Venkatesh: So here is a fantastic maths mark by a girl who is getting 97.

Professor Madhavan Mukund: So that is 233.

Professor G. Venkatesh: So here is a boy who is not doing that well, maths mark 44.

Professor Madhavan Mukund: So that is 271.

Professor G. Venkatesh: We can do like this we can go through the whole set I guess.

Professor Madhavan Mukund: So at the end we would have got some total.

Professor G. Venkatesh: We have got a total, so we have gone through the whole list now, what do we get?

Professor Madhavan Mukund: So I think the whole list there were 17 boys and 13 girls and the sum of the boys was 1220 and the sum of the girls was 951. So now what we wanted to do was find the average...

Professor G. Venkatesh: Average.

Professor Madhavan Mukund: So we need to divide, so if we actually divide the left hand side you get some 71.76 and this is some 73.15, so 1220 divided by 17, 71.76, 951 divided by 13 is 73.15.

Professor G. Venkatesh: So it looks like the girls are doing marginally better than the, not that much 73.15 is just a little bit more than 71.76, so the girls are doing better but one, could not

because it is small number 13 and so on you cannot really say much, but slightly marginally better than the boys.

Professor Madhavan Mukund: But what is interesting is that we were able to do this now in one scan, so we were able to iterate this set of cards once and keep track of these four outputs...

Professor G. Venkatesh: Four variables.

Professor Madhavan Mukund: As we were going along, so we were kind of doing a filtering on the gender in this case and we were accumulating both the number of cards of that gender and the maths marks total of that gender and finally we got all the totals and we were able to then...

Professor G. Venkatesh: So would you say that I mean we had this original question, are the girls doing better than boys, which seem like very interesting question, I mean it is kind of question that we normally ask but we have been able to turn that into a procedure, an iterator with some filtering and some variables and keeping track and all that and we have been able to give an answer which is numerically...

Professor Madhavan Mukund: Yes, that we can justify.

Professor G. Venkatesh: Justify, so would you say that this is an algorithm that we have been able to take this problem, are girls doing better than boys and we have found an algorithm for it, would you describe it like that?

Professor Madhavan Mukund: Yeah, that is a good way of thinking about it, so we have taken a question which is say subjective which you can debate about and we have given a systematic procedure on using the data available and a way to calculate and either confirm or to deny the hypothesis. So if we assume that girls are doing better than boys, then this is one way to validate whether this is true or not true, so that is true so this is very interesting that you can actually take a question which looks somewhat vague and make it precise by giving a criterion for that and giving an algorithm to compute that.

Professor G. Venkatesh: Okay.

Professor Madhavan Mukund: Very good.