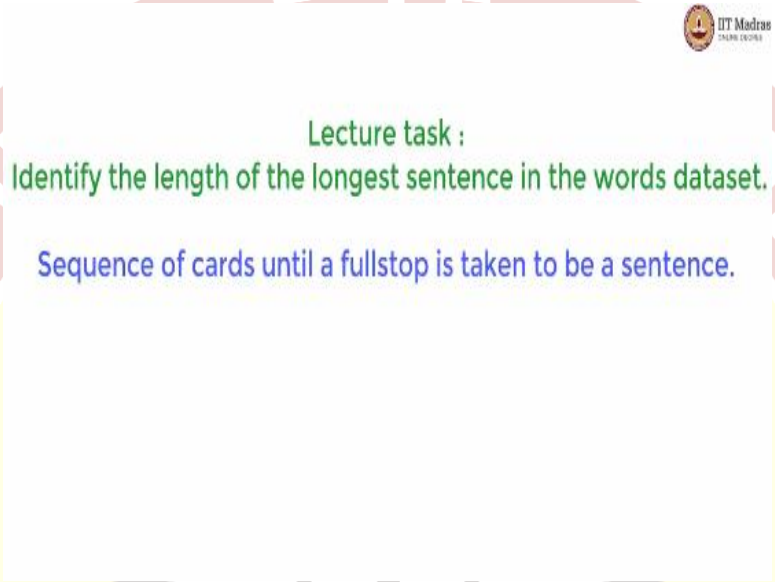
IIT Madras

ONLINE DEGREE

Computational Thinking
Indian Institute of Technology, Madras
Tutorial 2.2

Hello computational thinking students. This is the tutorial for lecture 12.

(Refer Slide Time: 00:18)



In lecture 12, professors found out the longest sentence in the words dataset. So, how did they identify a sentence while going through the iterations? They went card by card until a card contain a full stop. Full stop indicates the end of a sentence and the cards until that point until that full stop are taken to be one sentence. So, that number of cards is supposed to be a sentence.

(Refer Slide Time: 00:53)

So in order to identify the longest sentence, we first maintained a variable called longest sentence which was initialized to 0 and additionally we maintained another variable called count which is also initialize to 0 and count was updated for every card in a sentence. When a sentence was finished the longest sentence is compared against the count and if the count is greater that longest sentence, longest sentence was updated.

And then for the next sentence count was again reinitialized to 0 and card by card we count the number of words in that sentence. As you can see, when count is now greater than longest sentence, longest sentence is updated to the current count value. In this way at the end of iterating through the last card we have the longest sentence value to be equal to the number of cards in the longest sentence in the dataset. Now, let us attempt to do a similar task in a different dataset.

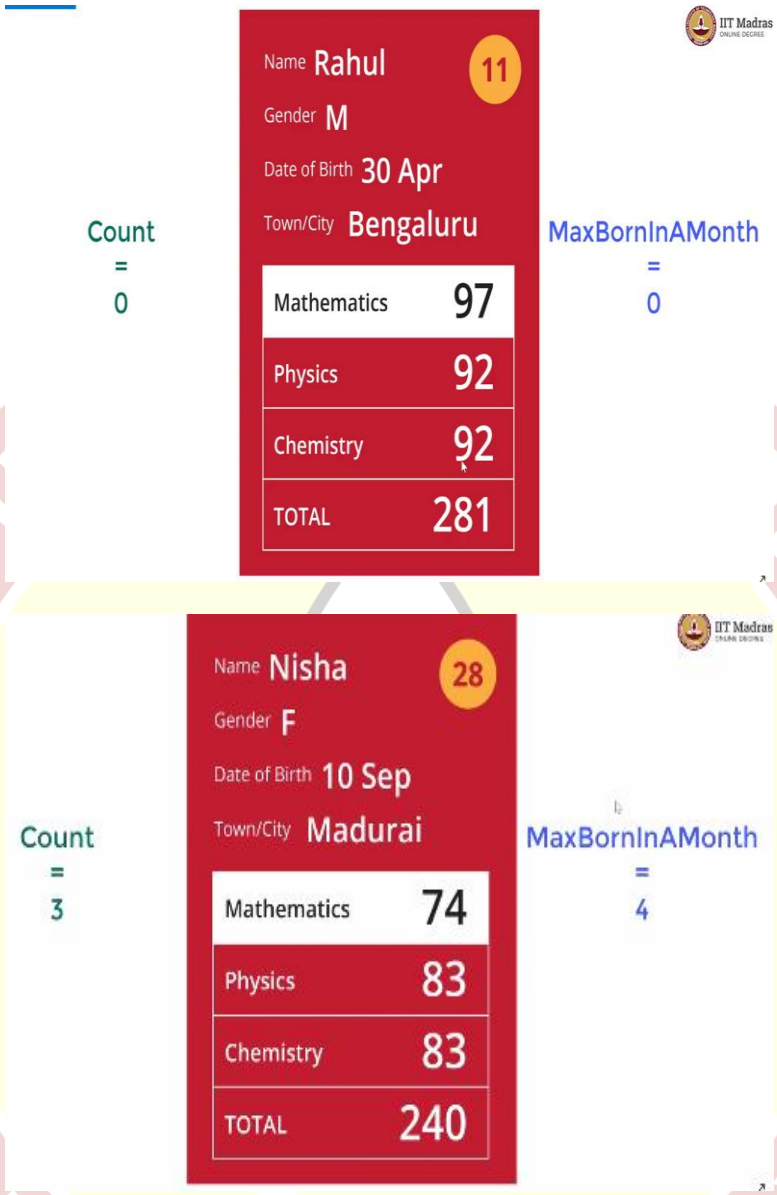
(Refer Slide Time: 02:20)

Let us take for example the scorecard dataset or you could even think about the shopping bill dataset and you notice that there is something fundamentally problematic with attempting this task on those datasets because the words dataset is a paragraph which is a set of ordered words. Each sentence is an ordered sequence of words and which is why when you go card after card you know you are in the same sentence until you see a full stop.

Whereas, in the words dataset or the shopping bill dataset the cards are not ordered in any way. So pulling out a sequence from these cards does not make much sense. So in order to be able to execute a similar task and the other datasets, what we could do is we could order them according to some property. For example, now if you look at this particular ordering of the scorecard dataset I have clubbed the cards of the same birth month together.

For example, this is April, this is August, this is December, December, December, December and then we move to Feb. So as you can see these cards are arranged according to the month of birth and so when go card after card you know you are in the same month until the month changes.

(Refer Slide Time: 04:07)



So very relevant task here would then be to find out what is the maximum number of students born in a particular month? So of all months, in one month what is the maximum number of students born? So, for this like the professors we first define a variable called max born in a month and we initialize it to 0 and then we need another variable which is the count variable which also we initialize to 0.

And this count we will keep incrementing for every card that we come across in a month and whenever the month changes we then update max in a month if it is lesser than the count. So we begin from the month of April the count variable is initialize to 0 and the max born in a month is also initialize to 0. The count variable is now incremented by 1 because we have found one card in April.

And we go to the next card and the month changes which means there has been only one card in the April month and thus the max born in a month variable needs to be updated to 1 because 0 is lesser than the 1 in count, and thus we have the max for in a month now updated to 1 and count is now reinitialize to 0 for the month of August and incremented by 1 for the one card that we have seen.

There is only one August card therefore no update in max because max is already one and the count is now reinitialize to 0 because the month has changed and incremented by 1 for the month of December. Now there is 2 December so count is 2, 3 December so the count is 3 and 4 December so the count is 4 and now the month is changed again which means our max variable needs to be updated to 4 because 4 is greater than 1.

And once that is done the count is now reinitialize to 0 and incremented by 1 for the month of February. Count is 2 and the month is changed. There is no update in max because 4 is greater than 2, and now count is reinitialize to 0 and incremented by 1 for the month of January. Count is 2, 3 and the month changes. So no update in the max variable count is reinitialize to 0 and incremented by 1 for the month of July.

That is 2 July cards, 3 July cards and the month is changed again no update in max. Count is reinitialize to 0 and incremented by 1 for the month of June. Only one June card, count is reinitialized to 0 and incremented to 1 for the month of March there has been no update in max because 4 is greater than 1. Now, count is 2, count is 3, count is 4 and the month changes. So max does not have to be updated because it is already at 4 and 4 is equal to 4.

So we now we initialize count to 0 and incremented by 1 for the month of May. 2, 3, 4 and again month changes which means max does not have to be updated. Count is reinitialized to 0 incremented by 1 for the month of November. This is 2 and November is done so max is not updated and count is reinitialized to 0 and incremented by 1 for the month of October. There is 2 October cards and the month is changed again.

Again, no update in max, count is reinitialize to 0 and incremented by 1 for the month of September, 2 September, 3 September and this is the last card. There is no card after this which means we do not need to update max, max is been at 4 and that is the maximum number of students born in a particular month. Thank you.