
# 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**
**Introduction to Datasets**

So, in our first lecture we talked about the kinds of questions that we are going to address in this course. So, we are going to do these patterns of computations and illustrate how they can be used to systematically calculate certain quantities and apply these patterns across different problems. So, for this we are going to use some specific types of data to make it more understandable and concrete.

(Refer Slide Time: 00:45)


So, first we will start with school marks. So, this is a typical report card, so it has various piece of information over it as name of the student, say the school, it also has some information about say the date of birth and various other factors to identify the student, it has a serial number for the student or the roll number. And then you have the marks in the different subjects in each subject maths, science, social studies, language and so on. So, there is a lot of information and this is a little bit complicated to process. So, what we will actually work with is a simplified form of these report cards.

(Refer Slide Time: 01:20)



This is the format in which we will look at the report card, so each report card has various pieces of information on it. It has a name, it has a gender, it has a date of birth in terms of a day and the month, a town or a city and then there are three marks, marks in maths, marks in physics, marks in chemistry and at the bottom there is a total which is a sum of these three marks.

And each card has a number on it to indicate it uniquely, so that if we happen to have say two people with the same name or something we would not get confused. So, this is card number 9, similarly this is card number 3 and this is card number 24 and so on. So, each card has exactly the same type of information but the exact details will vary.

So, this is for Siddharth and this is for Rida and this is a boy and this is a girl, this person was born in Madurai and this person was born in Chennai and so on. So, each report card has a certain amount of information on it and what we are going to do is try and understand how to calculate various quantities about this information.

(Refer Slide Time: 02:25)


So, our second set of data consist of shopping bills. So, if you look at the shopping bill typically it has a date, it has a name of the shop, it might have some other information about the shop like the address may be the GST number and various details which are not very important for us. And of course, it will have a list of items which have been bought.

So, it will have the name of the item the quantity, how many of that item were bought, it will have a unit price usually, how much each item cost if you buy one of them and then we will have a total for instance, if you buy 4 bananas and each banana cost 8 rupees then the total

will say 8 into 4, 32. And at the bottom of the bill you will have a total for the whole bill and then you might have some extra information like how much of tax you have paid, GST and so on. So, again this tax and all this extra thing is kind of extraneous and not so important for us.

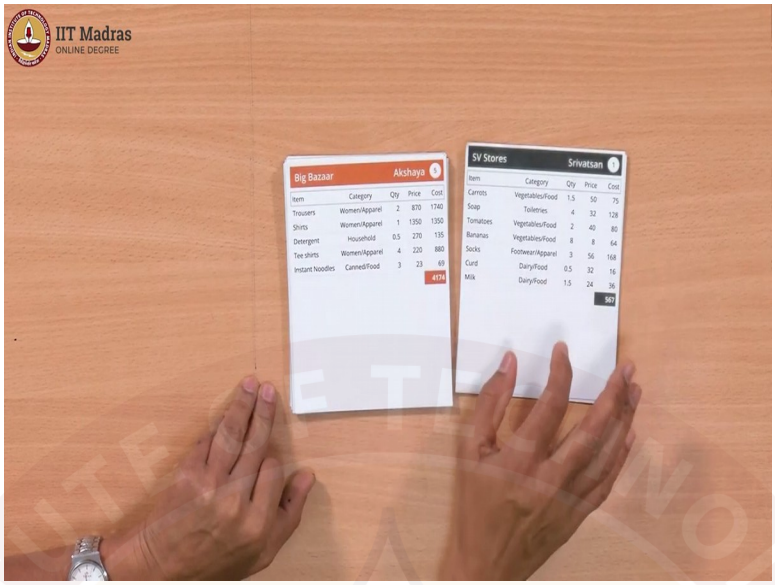
(Refer Slide Time: 03:21)

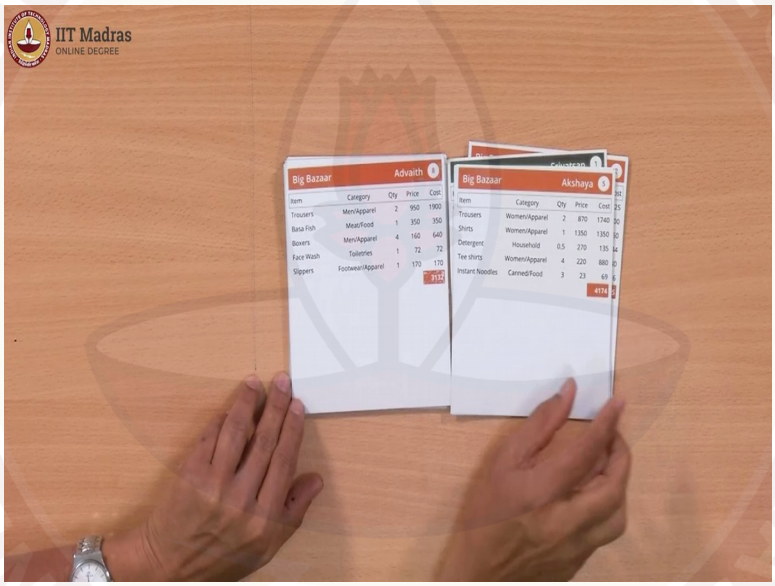

So, we will use a simplified form of the shopping bills. So, here is the typical shopping bill that we will use, so it has on the top the name of the shop here it is a SV stores, we have something which we will using for our calculation which we normally do not find on a shopping bill that is the name of the customer.

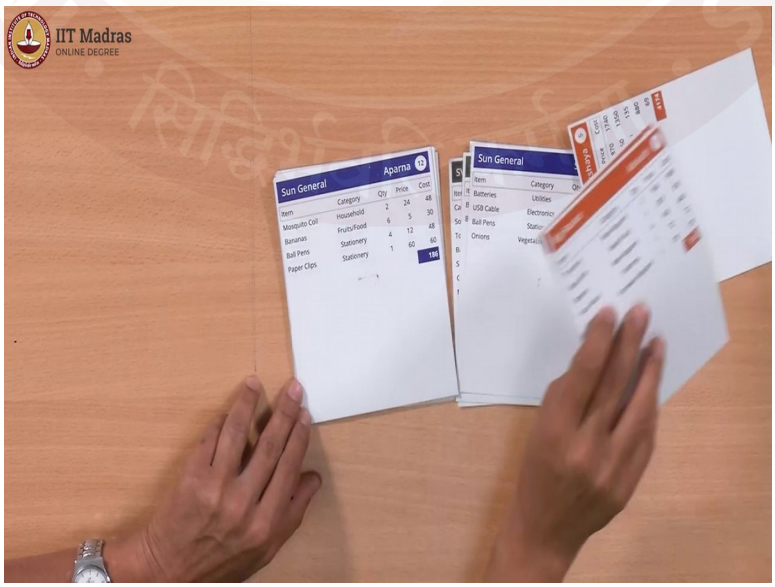
So, here it says, that this bill was Srivatsan's bill. Again it has a serial number 1 to indicate the this is the separate bill and the other ones will have numbers like 2, 3, 4 and so on. And then you have the details of course you have the items; carrots, soap, tomatoes, bananas, so these are the things that the person has bought in this bill.

What type of item it is, it is food or it is toiletry or is it footwear and so on. Then we have the quantity so this person has got 4 soaps, under the category toiletries each soap cost 32 rupees and the last price is, last column is the total cost the 32 rupees into 4 is 128 rupees. So, therefore each row now has 5 quantities, 5 items, the name of the item, the type, the quantity, the unit price and the total cost. And then if you come down at the bottom you have a grand total for this bill.

(Refer Slide Time: 04:23)

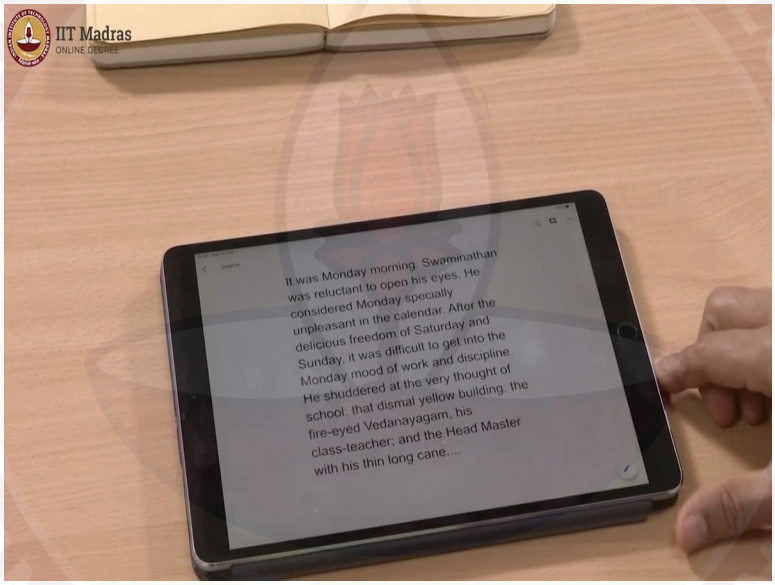


So, in this way we have several bills from different shops, so this is a shop Big Bazaar for Akshaya, this is another Big Bazaar and so on. Here is a third type of shop Sun general. So, what we will ask is questions about across shops, which shops are doing better business, which customers are more bigger spenders, who spending more on say food, who spending more on apparel, so these are all questions that we might want to ask and a shop might want to ask. Who are the customers who are shopping in all types of shops, who are the customers who are shopping in only one type of shop. So, there are many interesting questions you can ask about these bills and we will address some of them as we go along.

(Refer Slide Time: 05:08)




The third dataset that we will use is a collection of words taken from a paragraph. So, this paragraph is from Swami and friends, by the well-known author R.K Narayan. So, here is the paragraph. So what we have done with that paragraph is to put each word on a separate card. So, for instance the first word in the paragraph is ‘it’.

(Refer Slide Time: 05:21)


So, the card says the word, it has a serial number and it has two other items, it has the type of speech. So, 'it' is a pronoun and it also has the number of letters in the word. So, 'it' has two letters. So, each word is now represented, in each word in the paragraph is now represented as a separate card. And of course, here now it is important that you have the sequence number because some words like 'it' or 'was' will appear many times in the paragraph. So, we want to count each of them separately and make sure that they are separate cards.

(Refer Slide Time: 05:49)


So for example if you remember the paragraph is started off with 'it was Monday morning.' So, these are the first 4 cards. 'Swaminathan was reluctant to open his eyes.' So, this is just a systematic way in which we have taken the paragraph and broken it up in to single words. So,

we have taken the punctuation for example the full stop and put it along with a word, so that is not separated out.

And what we can now do is ask questions about these words, for instance, how many nouns are there, how many verbs are there, which is a longest word, how many sentences are there in which case we have to figure out how to decide when a sentence ends and so on. So, this will be our third dataset on which we will illustrate our computational thinking patterns.