

Week 1 : Variables , ,Initialization , Initialization , Iterators , Filtering , Datatypes , Flowcharts , Sanity of Data

Lecture 1 : Introduction To Dataset

moved to, the definition of *size* and *small*, everything else looks exactly the same across the three problems.

1.2 Sample datasets

We have picked some commonly used datasets that are simple enough to represent, but yet have sufficient structure to illustrate all the concepts that we wish to introduce through this book. The data could be represented using physical cards if the suggested procedures are to be carried out by hand as an activity in a classroom. We have chosen to use a tabular representation of the datasets for the book, since this is more compact and saves precious pages. There is a direct correspondence between a row in the table and a card, and we can work with table rows much as we do with physical cards. The accompanying digital companion will make this explicit. It will use the tabular representation, with the key difference being that individual table rows are separate objects like cards that can be moved around at will.

1.2.1 Classroom dataset

A data element in this dataset is the marks card of one student, shown below:

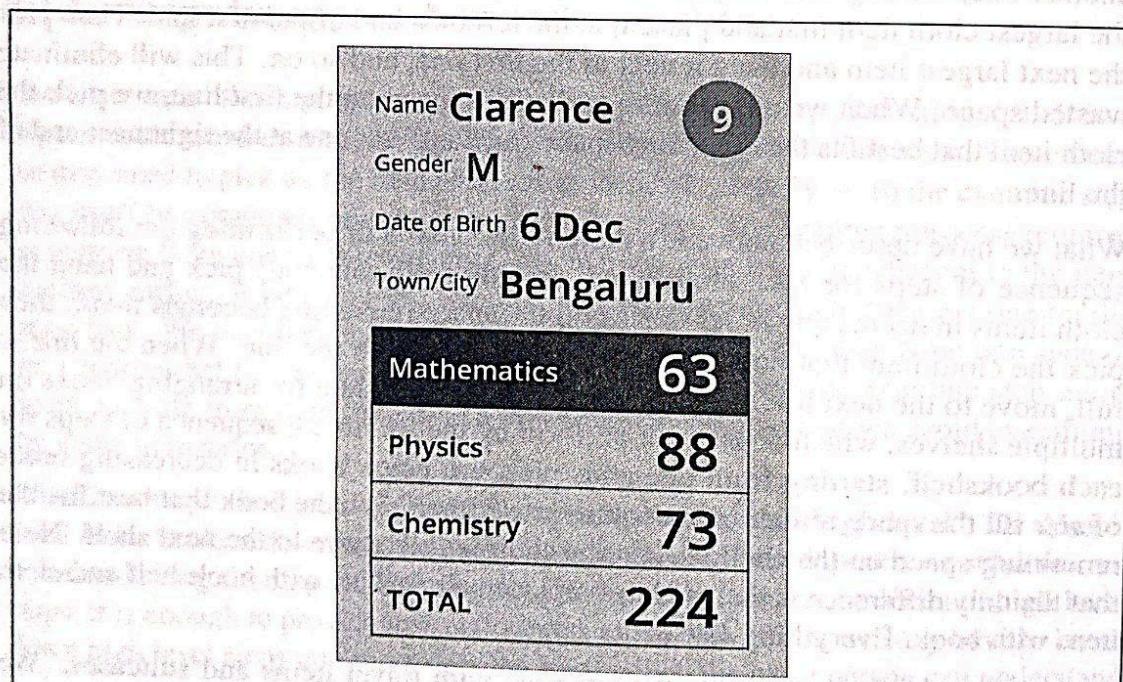


Figure 1.1: Grade card data element

- Unique id no: value 9 in the card above
- Name of student
- Gender: M for Male and F for Female
- Date of Birth: Only the day and month are recorded since this is sufficient for the questions that we will be asking related to this field
- Town/City
- Marks in Mathematics, Physics, Chemistry and the Total of these three marks

The dataset has 30 such cards with a mixture of Male and Female students from different cities.

Examples of questions that we can ask are:

- Which students are doing well, in each subject, and overall?
- Are girls doing better than boys?
- Can we group students based on their performance?
- Are there two students born on the same day and month?
- Are those scoring high marks in Maths also scoring high marks in Physics?
- Can we make pairs (A,B) of students in such a way that A can help B in one subject, while B can help A in another subject?

The entire data set is contained in the following table:

No	Name	Gen	DoB	Town/City	MA	PH	CH	Tot
0	Bhuvanesh	M	7 Nov	Erode	68	64	78	210
1	Harish	M	3 Jun	Salem	62	45	91	198
2	Shashank	M	4 Jan	Chennai	57	54	77	188
3	Rida	F	5 May	Chennai	42	53	78	173
4	Ritika	F	17 Nov	Madurai	87	64	89	240
5	Akshaya	F	8 Feb	Chennai	71	92	84	247
6	Sameer	M	23 Mar	Ambur	81	82	87	250
7	Aditya	M	15 Mar	Vellore	84	92	76	252
8	Surya	M	28 Feb	Bengaluru	74	64	51	189
9	Clarence	M	6 Dec	Bengaluru	63	88	73	224
10	Kavya	F	12 Jan	Chennai	64	72	68	204
11	Rahul	M	30 Apr	Bengaluru	97	92	92	281
12	Srinidhi	F	14 Jan	Chennai	52	64	71	187
13	Gopi	M	6 May	Madurai	65	73	89	227
14	Sophia	F	23 July	Trichy	89	62	93	244
15	Goutami	F	22 Sep	Theni	76	58	90	224
16	Tauseef	M	30 Dec	Trichy	87	86	43	216
17	Arshad	M	14 Dec	Chennai	62	81	67	210
18	Abirami	F	9 Oct	Erode	72	92	97	261
19	Vetrivel	M	30 Aug	Trichy	56	78	62	196
20	Kalyan	M	17 Sep	Vellore	93	68	91	252
21	Monika	F	15 Mar	Bengaluru	78	69	74	221
22	Priya	F	17 Jul	Nagercoil	62	62	57	181
23	Deepika	F	13 May	Bengaluru	97	91	88	276
24	Siddharth	M	26 Dec	Madurai	44	72	58	174
25	Geeta	F	16 May	Chennai	87	75	92	254
26	JK	M	22 Jul	Chennai	74	71	82	227

27	Jagan	M	4 Mar	Madurai	81	76	52	209
28	Nisha	F	10 Sep	Madurai	74	83	83	240
29	Naveen	M	13 Oct	Vellore	72	66	81	219

Figure 1.2: Classroom dataset

1.2.2 Shopping bill dataset

The basic data element in this dataset is the shopping bill generated when a customer visits a shop and buys something, as shown below:

SV Stores		Srivatsan 1		
Item	Category	Qty	Price	Cost
Carrots	Vegetables/Food	1.5	50	75
Soap	Toiletries	4	32	128
Tomatoes	Vegetables/Food	2	40	80
Bananas	Vegetables/Food	8	8	64
Socks	Footwear/Apparel	3	56	168
Curd	Dairy/Food	0.5	32	16
Milk	Dairy/Food	1.5	24	36
				567

Figure 1.3: Shopping bill data element

The card carries the following fields:

- Unique id no: value 1 in the card above
- Name of the store: SV Stores
- Name of the customer: Srivatsan
- List of items purchased: Item name, category name/sub category name, Quantity purchased, Price per unit, Cost of item purchased
- Total bill value

Several other fields normally present in shopping bills are omitted. For instance, tax amounts or percentages (GST), store address, phone number etc. These are not required for the questions we are going to ask about these data elements. The dataset has 30 such cards (bills) with a mixture of shops, customers and item categories.

- Are the high spenders shopping at shops with higher revenue?
- Which category is selling more?
- Which shop attracts the most number of loyal customers (those who go only to one shop)?
- Which customers are similar to each other in their shopping behaviour?

The entire data set is contained in the following table:

BNo	Shop	Customer				
1	SV Stores	Srivatsan				
	Item	Category	Qty	Price	Cost	Total
	Carrots	Vegetables/Food	1.5	50	75	567
	Soap	Toiletries	4	32	128	
	Tomatoes	Vegetables/Food	2	40	80	
	Bananas	Vegetables/Food	8	8	64	
	Socks	Footwear/Apparel	3	56	168	
	Curd	Dairy/Food	0.5	32	16	
	Milk	Dairy/Food	1.5	24	36	
2	Big Bazaar	Sudeep				
	Item	Category	Qty	Price	Cost	Total
	Baked Beans	Canned/Food	1	125	125	1525
	Chicken Wings	Meat/Food	0.5	600	300	
	Cocoa powder	Canned/Food	1	160	160	
	Capsicum	Vegetables/Food	0.8	180	144	
	Tie	Apparel	2	390	780	
	Clips	Household	0.5	32	16	
3	Sun General	Srivatsan				
	Item	Category	Qty	Price	Cost	Total
	Batteries	Utilities	6	14	84	354
	USB Cable	Electronics	1	85	85	
	Ball Pens	Stationery	5	12	60	
	Onions	Vegetables/Food	1.25	100	125	
4	SV Stores	Akshaya				
	Item	Category	Qty	Price	Cost	Total
	Face Wash	Toiletries	1	89	89	1341
	Shampoo	Toiletries	1	140	140	
	Onions	Vegetables/Food	1	98	98	
	Bananas	Fruits/Food	4	8	32	
	Milk	Dairy/Food	1	24	24	
	Biscuits	Packed/Food	2	22	44	
	Maggi	Packed/Food	1	85	85	
	Horlicks	Packed/Food	1	270	270	
	Chips	Packed/Food	1	20	20	
	Chocolates	Packed/Food	4	10	40	
	Cereal	Packed/Food	1	220	220	
	Handwash	Toiletries	1	139	139	
	Air freshener	Toiletries	2	70	140	

Chapter 1. Introduction

20

BNo	Shop	Customer				
			Qty	Price	Cost	Total
5	Big Bazaar	Akshaya				
	Item	Category				
	Trousers	Women/Apparel	2	870	1740	4174
	Shirts	Women/Apparel	1	1350	1350	
	Detergent	Household	0.5	270	135	
	Tee shirts	Women/Apparel	4	220	880	
	Instant Noodles	Canned/Food	3	23	69	
6	Sun General	Rajesh				
	Item	Category				
	Notebooks	Stationery	3	20	60	378
	Apples	Fruits/Food	6	24	144	
	Pears	Fruits/Food	4	30	120	
	Chart Paper	Stationery	2	22	44	
	Ruler	Stationery	1	10	10	
7	SV Stores	Advaith				
	Item	Category				
	Milk	Dairy/Food	2	24	48	123
	Bread	Packed/Food	1	30	30	
	Eggs	Food	1	45	45	
8	Big Bazaar	Advaith				
	Item	Category				
	Trousers	Men/Apparel	2	950	1900	3132
	Basa Fish	Meat/Food	1	350	350	
	Boxers	Men/Apparel	4	160	640	
	Face Wash	Toiletries	1	72	72	
	Slippers	Footwear/Apparel	1	170	170	
9	Sun General	Aparna				
	Item	Category				
	Mosquito Coil	Household	2	24	48	186
	Bananas	Fruits/Food	6	5	30	
	Ball Pens	Stationery	4	12	48	
	Paper Clips	Stationery	1	60	60	
10	SV Stores	Akhil				
	Item	Category				
	Bread	Packed/Food	1	30	30	96
	Biscuits	Packed/Food	3	22	66	
11	Big Bazaar	Mohith				
	Item	Category				
	Lindt	Chocolate/Food	1	125	125	595
	Socks	Footwear/Apparel	1	120	120	

		Spring Onions	Vegetables/Food	0.5	220	110	
		Lettuce	Vegetables/Food	0.6	150	90	
		Cookies	Snacks/Food	2	75	150	
BNo	Shop	Customer					
12	Sun General	Vignesh					
	Item	Category	Qty	Price	Cost	Total	
	Phone Charger	Utilities	1	230	230	656	
	Razor Blades	Grooming	1	12	12		
	Razor	Grooming	1	45	45		
	Shaving Lotion	Grooming	0.8	180	144		
	Earphones	Electronics	1	210	210		
	Pencils	Stationery	3	5	15		
BNo	Shop	Customer					
13	SV Stores	Abhinav					
	Item	Category	Qty	Price	Cost	Total	
	Chocolates	Packed/Food	1	10	10	893	
	Cereal	Packed/Food	1	220	220		
	Bananas	Fruits/Food	6	8	48		
	Tomatoes	Vegetables/Food	1	40	40		
	Curd	Dairy/Food	1	32	32		
	Milk	Dairy/Food	2	24	48		
	Horlicks	Packed/Food	1	270	270		
	Plates	Household	4	45	180		
	Eggs	Food	1	45	45		
BNo	Shop	Customer					
14	Big Bazaar	Abhinav					
	Item	Category	Qty	Price	Cost	Total	
	Shoes	Footwear/Apparel	1	2700	2700	3060	
	Polish	Footwear/Apparel	1	120	120		
	Socks	Footwear/Apparel	2	120	240		
BNo	Shop	Customer					
15	Sun General	Abhinav					
	Item	Category	Qty	Price	Cost	Total	
	Keyboard	Electronics	1	780	780	1100	
	Mouse	Electronics	1	320	320		
BNo	Shop	Customer					
16	SV Stores	George					
	Item	Category	Qty	Price	Cost	Total	
	Cereal	Packed/Food	1	220	220	279	
	Milk	Dairy/Food	1	24	24		
	Cupcakes	Packed/Food	1	25	25		
	Chocolates	Packed/Food	1	10	10		
BNo	Shop	Customer					
17	Big Bazaar	Radha					
	Item	Category	Qty	Price	Cost	Total	
	Broccoli	Vegetables/Food	0.5	120	60	798	
	Chicken Legs	Meat/Food	0.5	320	160		
	Basa Fish	Meat/Food	1	350	350		

		Lettuce Eggs	Vegetables/Food Meat/Food	0.8 12	150 9	120 108	
BNo	Shop	Customer					
18	Sun General	Advaith					
	Item	Category	Qty	Price	Cost	Total	
	Pencils	Stationery	2	5	10	187	
	Notebooks	Stationery	4	20	80		
	Geometry Box	Stationery	1	72	72		
	Graph Book	Stationery	1	25	25		
BNo	Shop	Customer					
19	SV Stores	Ahmed					
	Item	Category	Qty	Price	Cost	Total	
	Tomatoes	Vegetables/Food	1	40	40	603	
	Curd	Dairy/Food	1	32	32		
	Cupcakes	Packed/Food	2	25	50		
	Carrots	Vegetables/Food	1.5	50	75		
	Beans	Vegetables/Food	1	45	45		
	Onions	Vegetables/Food	0.5	98	49		
	Turmeric	Packed/Food	1	82	82		
	Ghee	Packed/Food	1	230	230		
BNo	Shop	Customer					
20	Sun General	Suresh					
	Item	Category	Qty	Price	Cost	Total	
	Batteries	Utilities	2	14	28	315	
	Tomatoes	Vegetables/Food	1.5	80	120		
	Spinach	Vegetables/Food	1	15	15		
	Bananas	Fruits/Food	4	5	20		
	Mosquito coils	Household	1	24	24		
	Guava	Fruits/Food	0.4	120	48		
	Potato	Vegetables/Food	1.5	40	60		
BNo	Shop	Customer					
21	SV Stores	Ahmed					
	Item	Category	Qty	Price	Cost	Total	
	Tomatoes	Vegetables/Food	2	40	80	622	
	Curd	Dairy/Food	2	32	64		
	Cupcakes	Packed/Food	3	25	75		
	Carrots	Vegetables/Food	0.5	50	25		
	Onions	Vegetables/Food	1	98	98		
	Handwash	Toiletries	1	139	139		
	Bananas	Fruits/Food	12	8	96		
	Eggs	Food	1	45	45		
BNo	Shop	Customer					
22	SV Stores	Neeraja					
	Item	Category	Qty	Price	Cost	Total	
	Carrots	Vegetables/Food	1	50	50	592	
	Horlicks	Packed/Food	1	270	270		
	Chips	Packed/Food	1	20	20		
	Kajal	Cosmetics	1	180	180		
	Milk	Dairy/Food	3	24	72		

BNo	Shop	Customer	Category	Qty	Price	Cost	Total
23	SV Stores	Akshaya	Curd	0.5	32	16	128
			Butter	0.2	320	64	
			Milk	2	24	48	
24	SV Stores	Julia	Carrots	1.5	50	75	888
			Bananas	12	8	96	
			Curd	3	32	96	
			Milk	4	24	96	
			Cereal	2	220	440	
			Maggi	1	85	85	
25	Sun General	Ahmed	Earphones	1	210	210	1364
			Phone cover	1	140	140	
			Dongle	1	790	790	
			A4 sheets	200	1	200	
			Ball Pens	2	12	24	
26	SV Stores	Srivatsan	Beans	1	45	45	276
			Bread	1	30	30	
			Onions	0.5	98	49	
			Bananas	6	8	48	
			Curd	1	32	32	
			Milk	3	24	72	
27	Sun General	Srivatsan	Broom	1	70	70	340
			Dustpan	1	45	45	
			Floor Cleaner	1	125	125	
			Tissue Paper	2	50	100	
28	SV Stores	Neeraja	Milk	3	24	72	92
			Chips	1	20	20	
29	SV Stores	Vignesh	Item	Category	Qty	Price	Cost Total

5

Face Wash	Toiletries	1	89	89	514
Shampoo	Toiletries	1	140	140	
Maggi	Packed/Food	1	85	85	
Chips	Packed/Food	1	20	20	
Chocolates	Packed/Food	4	10	40	
Air Freshener	Toiletries	2	70	140	
BNo	Shop	Customer			
30	SV Stores	Ahmed			
	Item	Category	Qty	Price	Cost
	Chocolates	Packed/Food	1	10	10
	Curd	Dairy/Food	2	32	64
	Bananas	Fruits/Food	4	8	32
					Total
					106

Figure 1.4: Shopping bill dataset

1.2.3 Words dataset

The basic data element in this dataset, shown below, is one word taken from a paragraph of text from the book *Swami and Friends* by R K Narayanan. The first sentence, consisting of 4 words, is shown below:

0	It	1	was
Pronoun	Letter count: 2	Verb	Letter count: 3
2	Monday	3	morning.
Noun	Letter count: 6	Noun	Letter count: 7

Figure 1.5: Words data elements

The card carries the following fields:

- Unique id no: values 0 to 3 in the cards above
- The word itself
- Part of speech or word category: Pronoun, Verb, Noun for the cards above

- Letter count: 2, 3, 6 and 7 respectively for the cards above

The last word in the sentence also carries a full stop. Punctuation marks are stored along with corresponding words. Note that the full stop (and other punctuation mark) are not included in the letter count. The part of speech and letter count characteristics are similar to what you may find in a dictionary entry for that word. The dataset has 65 such cards (words) that vary with respect to parts of speech and number of letters.

Examples of questions that we can ask are:

- How many sentences does the paragraph have?
- Which words occur with high frequency?
- Are higher frequency words shorter?
- How does one resolve the pronoun to a personal noun?
- Is the paragraph coherent (i.e., talking about a connected set of nouns)?

The entire data set is contained in the following table:

Sno	Word	Part of Speech	Letter count
0	It	Pronoun	2
1	was	Verb	3
2	Monday	Noun	6
3	morning.	Noun	7
4	Swaminathan	Noun	11
5	was	Verb	3
6	reluctant	Adjective	9
7	to	Preposition	2
8	open	Verb	4
9	his	Pronoun	3
10	eyes.	Noun	4
11	He	Pronoun	2
12	considered	Verb	10
13	Monday	Noun	6
14	specially	Adverb	9
15	unpleasant	Adjective	10
16	in	Preposition	2
17	the	Article	3
18	calendar.	Noun	8
19	After	Preposition	5
20	the	Article	3
21	delicious	Adjective	9
22	freedom	Noun	7
23	of	Preposition	2
24	Saturday	Noun	8
25	And	Conjunction	3
26	Sunday,	Noun	6
27	it	Pronoun	2
28	was	Verb	3
29	difficult	Adjective	9
30	to	Preposition	2
31	get	Verb	3
32	into	Preposition	4
33	the	Article	3

Chapter 1. Introduction

34	Monday	Noun	6
35	mood	Noun	4
36	of	Preposition	2
37	work	Noun	4
38	and	Conjunction	3
39	discipline.	Noun	10
40	He	Pronoun	2
41	shuddered	Verb	9
42	at	Preposition	2
43	the	Article	3
44	very	Adverb	4
45	thought	Noun	7
46	of	Preposition	2
47	school:	Noun	6
48	the	Article	3
49	dismal	Adjective	6
50	yellow	Adjective	6
51	building;	Noun	8
52	the	Article	3
53	fire-eyed	Adjective	8
54	Vedanayagam,	Noun	11
55	his	Pronoun	3
56	class	Noun	5
57	teacher,	Noun	7
58	and	Conjunction	3
59	headmaster	Noun	10
60	with	Preposition	4
61	his	Pronoun	3
62	thin	Adjective	4
63	long	Adjective	4
64	cane ...	Noun	4

Figure 1.6: Words dataset

1.2.4 Train timetable dataset

There are two basic data elements in this dataset — a station and a train, shown below.

Station Card (Nagpur):

Train No.	Arr.	Dep.	Days
12281	04:10	04:15	M W Th F
12221	04:10	04:15	Tu Su
02286	05:15	05:20	Tu Su
12270	05:15	05:20	W Su
12289	07:20	--	M Tu W Th F Sa Su
12290	--	20:40	M Tu W Th F Sa Su
12269	21:00	21:00	M F
02285	21:00	21:00	Th Su
12222	23:15	23:20	Th Sa
12262	23:15	23:20	M Tu W F

Train Card (Train 02286):

Dist.	Station Name	Arr.	Dep.	Day
0	New Delhi	--	15:45	1
403	Jhansi	20:22	20:27	1
700	Bhopal	00:01	00:06	2
1083	Nagpur	05:15	05:20	2
1294	Balharshah	08:10	08:15	2
1660	Secunderabad	14:00	--	2

Figure 1.7: Station and Train data elements

The station card carries the following fields:

- Name of station: Nagpur
- List of trains stopping at the station: Train Number, Arrival Time, Departure Time, List of days of the week when the train is running

Note that the days of the week represents the days when the train stops at Nagpur. The train may have started from its origin station on an earlier day. So for example, 12289 and 12290 run on all days of the week, while 12221 will arrive/depart at Nagpur only on Tuesdays and Sundays.

A blank entry in the arrival time field for a train (e.g. 12290) indicates that the train originates at Nagpur. A blank entry in the departure time field for a train (e.g. 12289) indicates that the train terminates at Nagpur.

The train card carries the following fields:

- Number of the train: 02286
- List of days of the week that the train runs: M,F
- List of stations at which the train stops: Distance from origin station in km, Station name, Arrival Time, Departure Time, Day from start of train

Note that the day from the start shows 1 if it is the first day (i.e., same date as the start of the train), and will show 2 for the day after the start date, 3 if it is two days after the start date, etc. As with stations, a blank entry in the arrival time field for a station means that station is the origin station for the train, with a distance of 0, and a blank entry in the departure time field means that the station is the terminal station for the train.

Note also that the stations refer to the trains, and the trains refer to the stations!

Examples of questions that we can ask are:

5

- Is there a direct train between two given stations?
- What is the distance travelled on a given train between two stations?
- Can I go from a given station to another by changing at most one train?
- How many stations can I reach from a given station if I start today and am willing to travel for two full days?
- What is the shortest distance between two given stations?
- If I am at station A at a given time T, then what is the earliest time that I can reach another station B?
- Is it possible to go from one station to another and come back on the same day?
- Which stations can you call junctions?
- Which is the busiest junction?
- What is the best way to visit a list of stations and come back to my home location?

6

The entire station data set is contained in the following table:

Sno	Station Name	Train No.	Arr.	Dep.	Running days
1	Adra	12282	03:40	03:50	F
		12281	14:56	15:00	W
Sno	Station Name	Train No.	Arr.	Dep.	Running days
2	Ahmedabad	12267	05:55	06:15	M Tu W Th F Sa Su
		12298	06:30	—	Tu F Su
		12268	23:15	23:36	M Tu W Th F Sa Su
		12297	—	22:25	M Th Su
Sno	Station Name	Train No.	Arr.	Dep.	Running days
3	Asansol	02213	01:08	01:12	Tu Th Sa
		02214	01:56	02:01	W F Su
		12274	07:35	07:40	W Su
		12273	10:56	11:00	M F
Sno	Station Name	Train No.	Arr.	Dep.	Running days
4	Balasore	12282	07:33	07:38	F
		12281	10:44	10:49	W
Sno	Station Name	Train No.	Arr.	Dep.	Running days
5	Balharshah	02286	08:10	08:15	Tu Sa

12270	08:10	08:15	W Su
12213	13:31	13:35	Su
12214	16:56	17:00	Tu
12269	18:06	18:11	M F
02285	18:06	18:11	Th Su

Sno	Station Name	Train No.	Arr.	Dep.	Running days
6	Bengaluru	12214	07:55	-	W
		02246	-	11:00	M Tu Th F Su
		02245	16:00	-	M W Th Sa Su
		12213	-	23:40	Sa

Sno	Station Name	Train No.	Arr.	Dep.	Running days
7	Bhopal	02286	00:01	00:06	Tu Sa
		02285	02:15	02:20	M F
		12269	02:18	02:20	Tu Sa
		12214	08:20	08:25	Tu
		12213	21:20	21:25	Su
		12270	23:59	00:04	W Su

Sno	Station Name	Train No.	Arr.	Dep.	Running days
8	Bhubaneswar	02201	06:30	-	Tu Th Sa
		12281	-	07:40	W
		02246	09:50	10:00	M Tu W F Sa
		12282	11:15	-	F
		02245	16:50	17:00	Tu W F Sa Su
		02202	-	20:15	Tu Th Sa

Sno	Station Name	Train No.	Arr.	Dep.	Running days
9	Bhusaval	12290	01:20	01:25	M Tu W Th F Sa Su
		12289	02:00	02:05	M Tu W Th F Sa Su
		12222	03:42	03:47	F Su
		12262	03:42	03:47	Tu W Th Sa
		12294	08:56	09:00	W Su
		12221	22:40	22:45	M Sa
		12261	23:00	23:06	Tu W Th Su
		12293	23:00	23:06	M F

Sno	Station Name	Train No.	Arr.	Dep.	Running days
10	Bikaner	12260	-	12:15	M Tu Th F
		12259	19:15	-	M Tu Th F

Sno	Station Name
11	Bilaspur

5

Train No.	Arr.	Dep.	Running days
12261	09:56	10:06	M W Th F
12221	09:56	10:06	Tu Su
12222	17:36	17:45	Th Sa
12262	17:36	17:45	M Tu W F

Sno	Station Name	Train No.	Arr.	Dep.	Running days
12	Chennai	12269	-	06:40	M F
		12270	20:40	-	W Su

Sno	Station Name	Train No.	Arr.	Dep.	Running days
13	Churu	12260	14:36	14:40	M Tu Th F
		12259	16:06	16:08	M Tu Th F

Sno	Station Name	Train No.	Arr.	Dep.	Running days
14	Daund	12222	10:00	10:15	F Su
		12221	16:20	16:36	M Sa

Sno	Station Name	Train No.	Arr.	Dep.	Running days
15	Dhanbad	12260	08:58	09:03	Tu W F Sa
		12259	21:52	21:56	M W Th Su

Sno	Station Name	Train No.	Arr.	Dep.	Running days
16	Ernakulam	02284	18:20	-	M
		12223	20:15	-	W Su
		02283	-	23:25	Tu
		12224	-	23:25	W Su

Sno	Station Name	Train No.	Arr.	Dep.	Running days
17	Guntakal	12214	02:58	03:00	W
		12213	03:45	03:50	Su

Sno	Station Name	Train No.	Arr.	Dep.	Running days
18	Gwalior	12269	06:51	06:53	Tu Sa
		12270	19:09	19:11	Tu Sa

Sno	Station Name	Train No.	Arr.	Dep.	Running days
19	Kolkata (1/2)				

02202	04:15	—	W F Su
02214	05:40	—	W F Su
12222	—	08:20	Th Sa
12262	—	08:20	M Tu W F
12273	—	08:35	M F
12274	10:40	—	W Su
02245	—	10:50	Tu W F Sa Su
12260	12:45	—	Tu W F Sa
02246	16:45	—	M Tu W F Sa
12259	—	18:31	M W Th Su
Sno	Station Name		
20	Kolkata (2/2)		
	Train No.	Arr.	Dep.
	12261	19:50	—
	12221	19:50	—
	02201	—	20:00
	02213	—	22:00
Sno	Station Name		
21	Igatpuri		
	Train No.	Arr.	Dep.
	12290	05:10	05:15
	12262	07:40	07:45
	12294	12:31	12:35
	12261	19:25	19:31
	12293	19:25	19:31
	12289	22:25	22:31
Sno	Station Name		
22	Indore		
	Train No.	Arr.	Dep.
	12227	11:06	—
	12228	—	23:00
Sno	Station Name		
23	Jabalpur		
	Train No.	Arr.	Dep.
	12294	00:35	00:45
	12293	05:50	06:00
Sno	Station Name		
24	Jaipur		
	Train No.	Arr.	Dep.
	12239	14:35	—
	12240	—	19:11
Sno	Station Name		
25	Jammu		
	Train No.	Arr.	Dep.
	12266	—	19:20
	12265	07:25	—
Sno	Station Name		

26 Jasidih

Train No.	Arr.	Dep.	Running days
12274	05:06	05:09	W Su
12273	12:25	12:27	M F

Sno 27 Station Name Jhansi

Train No.	Arr.	Dep.	Running days
12213	01:15	01:20	M
12214	04:20	04:25	Tu
12269	05:30	05:35	Tu Sa
02285	05:30	05:35	M F
02286	20:22	20:27	M F
12270	20:22	20:27	Tu Sa

Sno 28 Station Name Kanpur

Train No.	Arr.	Dep.	Running days
12260	00:29	00:34	Tu W F Sa
12273	00:40	00:45	Tu Sa
12281	00:50	00:55	Th
12259	06:25	06:30	M Tu Th F
12274	17:48	17:54	Tu Sa
12282	17:48	17:54	Th

Sno 29 Station Name Kharagpur

Train No.	Arr.	Dep.	Running days
02202	01:10	01:15	W F Su
02201	22:15	22:20	M W F

Sno 30 Station Name Kota

Train No.	Arr.	Dep.	Running days
12263	01:00	01:10	W Sa
02284	02:55	03:05	Su
22210	04:15	04:25	W Su
22209	10:25	10:31	Tu Sa
02283	12:50	12:56	Th
12264	15:20	15:31	M Th

Sno 31 Station Name Kozhikode

Train No.	Arr.	Dep.	Running days
02283	02:47	02:50	W
12224	02:47	02:50	M Th
12223	15:23	15:25	W Su
02284	13:27	13:31	M

Sno 32 Station Name Loharu

Train No.	Arr.	Dep.	Running days
12259	14:23	14:26	M Tu Th F

		12260	16:08	16:11	M Tu Th F
Sno 33	Station Name Lonavla	Train No.	Arr.	Dep.	Running days
		12264	06:04	06:08	Tu F
		12297	06:05	06:07	M Tu F
		12263	11:58	12:00	Tu F
		12298	22:18	22:20	M Th Sa
Sno 34	Station Name Ludhiana	Train No.	Arr.	Dep.	Running days
		12265	02:45	02:55	M W Sa
		12266	23:15	23:25	M W Sa
Sno 35	Station Name Madgaon	Train No.	Arr.	Dep.	Running days
		02284	04:25	04:35	M
		12223	07:10	07:20	W Su
		02283	12:50	13:00	W
		12224	12:50	13:00	M Th
Sno 36	Station Name Mangalore	Train No.	Arr.	Dep.	Running days
		02283	06:10	06:20	W
		12224	06:10	06:20	M Th
		02284	10:15	10:25	M
		12223	12:10	12:20	W Su
Sno 37	Station Name Manmad	Train No.	Arr.	Dep.	Running days
		12222	06:00	06:04	F Su
		12221	20:31	20:33	M Sa
Sno 38	Station Name Mumbai (1/3)	Train No.	Arr.	Dep.	Running days
		02283	00:00	00:10	Th
		12298	00:40	00:45	Tu F Su
		12264	03:50	03:55	Tu F
		12297	03:50	03:55	M Tu F
		12268	06:00	-	M Tu W Th F Sa Su
		12290	08:06	-	M Tu W Th F Sa Su
		12262	10:31	-	Tu W Th Sa
		12228	10:50	-	M Sa
		12240	10:50	-	W F
		12220	11:06	-	W Sa
Sno 39	Station Name Mumbai (2/3)				

Train No.	Arr.	Dep.	Running days
12263	14:35	14:40	Tu F
12294	14:56	—	W Su
22210	16:15	—	W Su
12261	—	17:15	Tu W Th Su
12293	—	17:25	M F
02284	17:56	18:00	Su
12289	—	20:15	M Tu W Th F Sa Su
12223	—	20:50	Tu Sa
02283	22:40	22:45	W
12219	—	23:06	W Sa

Sno	Station Name	Arr.	Dep.	Running days
40	Mumbai (3/3)			
	Train No.	Arr.	Dep.	Running days
	12227	—	23:15	Th Sa
	12239	—	23:15	Tu Su
	22209	—	23:15	M F
	12267	—	23:25	M Tu W Th F Sa Su
	12224	23:56	—	M Th

Sno	Station Name	Arr.	Dep.	Running days
41	Nagpur			
	Train No.	Arr.	Dep.	Running days
	12261	04:10	04:15	M W Th F
	12221	04:10	04:15	Tu Su
	02286	05:15	05:20	Tu Sa
	12270	05:15	05:20	W Su
	12289	07:20	—	M Tu W Th F Sa Su
	12290	—	20:40	M Tu W Th F Sa Su
	12269	21:00	21:06	M F
	02285	21:00	21:06	Th Su
	12222	23:15	23:20	Th Sa
	12262	23:15	23:20	M Tu W F

Sno	Station Name	Arr.	Dep.	Running days
42	New Delhi			
	Train No.	Arr.	Dep.	Running days
	12266	04:20	—	Tu Th Su
	12281	06:25	—	Th
	12273	06:25	—	Tu Sa
	12263	06:55	—	W Sa
	12213	07:00	—	M
	12269	10:40	—	Tu Sa
	02285	10:40	—	M F
	12264	—	10:56	M Th
	12259	11:31	11:45	M Tu Th F
	12274	—	12:56	Tu Sa

Sno	Station Name	Arr.	Dep.	Running days
43	New Delhi			
	Train No.	Arr.	Dep.	Running days
	02286	—	15:45	M F
	12270	—	15:45	Tu Sa

1.2 Sample datasets

35

22209	16:31	-	Tu Sa
12260	19:15	19:40	M Tu Th F
02283	19:40	-	Th
02284	-	21:36	Sa
12265	-	22:15	Tu F Sa
12214	-	23:00	M
22210	-	23:25	Tu Sa

Sno Station Name

44 Patna

Train No.	Arr.	Dep.	Running days
12274	01:10	01:20	W Sa
02213	06:40	-	Tu Th Sa
12273	16:20	16:31	M F
02214	-	20:40	Tu Th Sa

Sno Station Name

45 Prayagraj

Train No.	Arr.	Dep.	Running days
12293	12:45	-	Tu Sa
12294	-	19:20	Tu Sa

Sno Station Name

46 Punc

Train No.	Arr.	Dep.	Running days
12219	02:05	02:10	Th Sa
12264	07:10	-	Tu F
12297	07:10	-	M Tu F
12220	07:50	07:55	W Sa
12263	-	11:10	Tu F
12222	11:45	-	F Sa
12221	-	15:15	M Sa
12298	-	21:36	M Th Sa

Sno Station Name

47 Rajkot

Train No.	Arr.	Dep.	Running days
12268	-	19:06	M Tu W Th F Sa Su
12267	10:56	-	M Tu W Th F Sa Su

Sno Station Name

48 Ratangarh

Train No.	Arr.	Dep.	Running days
12260	13:50	13:52	M Tu Th F
12259	17:00	17:02	M Tu Th F

Sno Station Name

49 Ratlam

Train No.	Arr.	Dep.	Running days
12228	01:50	01:55	M Sa
12240	01:50	01:55	W F
02284	07:05	07:10	Su
12227	07:25	07:30	F Su
22210	07:30	07:35	W Su

12239	07:35	07:40	M W
22209	07:35	07:40	Tu Sa
02283	09:10	09:15	Th
12264	18:36	18:40	M Th
12263	22:11	22:15	Tu F

Sno	Station Name	Train No.	Arr.	Dep.	Running days
50	Ratnagiri	12223	02:50	02:55	W Su
		02283	17:11	17:15	W
		12224	17:11	17:15	M Th
		02284	23:20	23:25	Su

Sno	Station Name	Train No.	Arr.	Dep.	Running days
51	Renigunta	02245	10:10	10:15	M W Th Sa Su
		02246	16:50	17:00	M Tu Th F Su

Sno	Station Name	Train No.	Arr.	Dep.	Running days
52	Sadulpur	12259	15:03	15:06	M Tu Th F
		12260	15:25	15:27	M Tu Th F

Sno	Station Name	Train No.	Arr.	Dep.	Running days
53	Satna	12293	08:31	08:40	Tu Sa
		12294	22:00	22:11	Tu Sa

Sno	Station Name	Train No.	Arr.	Dep.	Running days
54	Sawai Madhopur	12239	12:10	12:20	M W
		12240	20:56	21:06	Tu Th

Sno	Station Name	Train No.	Arr.	Dep.	Running days
55	Secunderabad	22203	06:15	—	M W F
		12213	08:56	09:10	Su
		12219	11:06	—	Th Su
		02285	—	13:10	Th Su
		02286	14:00	—	Tu Sa
		22204	—	20:15	M W Sa
		12214	21:15	21:31	Tu
		12220	—	23:06	Tu F

Sno	Station Name	Train No.	Arr.	Dep.	Running days
56	Solapur				

		12220	03:57	04:00	W Sa
		12219	05:57	06:00	Th Su
Sno	Station Name				
57	Sri Dungargarah				
		Train No.	Arr.	Dep.	Running days
		12260	13:09	13:10	M Tu Th F
		12259	17:47	17:48	M Tu Th F
Sno	Station Name				
58	Surendranagar				
		Train No.	Arr.	Dep.	Running days
		12267	08:24	08:34	M Tu W Th F Sa Su
		12268	20:38	20:47	M Tu W Th F Sa Su
Sno	Station Name				
59	Tatanagar				
		Train No.	Arr.	Dep.	Running days
		12222	11:33	11:43	Th Sa
		12262	11:33	11:43	M Tu W F
		12261	15:56	16:06	M W Th F
		12221	15:56	16:06	Tu Su
Sno	Station Name				
60	Ujjain				
		Train No.	Arr.	Dep.	Running days
		12228	00:15	00:20	M Sa
		12227	08:50	08:56	F Su
Sno	Station Name				
61	Vadodara				
		Train No.	Arr.	Dep.	Running days
		12227	03:52	04:05	F Su
		12239	03:52	04:05	M W
		22209	03:52	04:05	Tu Sa
		02283	05:13	05:31	Th
		12228	05:20	05:30	M Sa
		12240	05:20	05:30	W F
		02284	11:01	11:20	Su
		22210	11:10	11:20	W Su
		12263	18:34	18:45	Tu F
		12264	22:29	22:38	M Th
Sno	Station Name				
62	Varanasi				
		Train No.	Arr.	Dep.	Running days
		12259	02:20	02:30	M Tu Th F
		12260	04:33	04:43	Tu W F Sa
		12273	20:13	20:23	M F
		12281	20:45	20:56	W
		12282	21:50	22:00	Th
		12274	22:11	22:20	Tu Sa
Sno	Station Name				

Chapter 1. Introduction

38

63	Vijayawada			
	Train No.	Arr.	Dep.	Running days
	22203	00:40	00:55	M W F
	22204	01:05	01:15	Tu Th Su
	02245	04:20	04:25	M W Th Sa Su
	12269	12:20	12:31	M F
	12270	14:20	14:25	W Su
	02246	22:20	22:31	M Tu Th F Su
64	Sno	Station Name		
		Visakhapatnam		
		Train No.	Arr.	Dep.
		22204	06:30	—
		22203	—	19:45
65	Sno	Station Name		
		Vizianagram		
		Train No.	Arr.	Dep.
		02246	04:30	04:40
				M Tu W F Sa

Figure 1.8: Station dataset

The entire train data set is contained in the following table:

Sno	Train No.	Running Days		
1	02202	Tu Th Sa		
	Dist.	Stn. Name	Arr.	Dep.
	0	Bhubaneswar	—	20:15
	387	Kharagpur	01:10	01:15
	522	Kolkata	04:15	—
Sno	Train No.	Running Days		
2	12281	W		
	Dist.	Stn. Name	Arr.	Dep.
	0	Bhubaneswar	—	07:40
	206	Balasore	10:44	10:49
	496	Adra	14:56	15:00
	964	Varanasi	20:45	20:56
	1311	Kanpur	00:50	00:55
	1750	New Delhi	06:25	—
Sno	Train No.	Running Days		
3	12260	M Tu Th F		
	Dist.	Stn. Name	Arr.	Dep.
	0	Bikaner	—	12:15
	74	Sri Dungargarh	13:09	13:10
	138	Ratangarh	13:50	13:52
	181	Churu	14:36	14:40
	238	Sadulpur	15:25	15:27

288	Loharu	16:08	16:11	1
463	New Delhi	19:15	19:40	1
902	Kanpur	00:29	00:34	2
1249	Varanasi	04:33	04:43	2
1651	Dhanbad	08:58	09:03	2
1916	Kolkata	12:45	-	2

Sno	Train No.	Running Days				
		Tu	W	Th	Su	
4	12261					
	Dist.					
	0	Mumbai	-	17:15	1	
	137	Igatpuri	19:25	19:31	1	
	445	Bhusaval	23:00	23:06	1	
	837	Nagpur	04:10	04:15	2	
	1250	Bilaspur	09:56	10:06	2	
	1719	Tatanagar	15:56	16:06	2	
	1968	Kolkata	19:50	-	2	

Sno	Train No.	Running Days				
		M	Tu	W	Th	Sa
5	12289	M	Tu	W	Th	Sa
	Dist.					
	0	Mumbai	-	20:15	1	
	137	Igatpuri	22:25	22:31	1	
	445	Bhusaval	02:00	02:05	2	
	837	Nagpur	07:20	-	2	

Sno	Train No.	Running Days				
		M				
6	12214					
	Dist.					
	0	Stn. Name	Arr.	Dep.	Day	
	423	New Delhi	-	23:00	1	
	720	Jhansi	04:20	04:25	2	
	1314	Bhopal	08:20	08:25	2	
	1680	Balharshah	16:56	17:00	2	
	2091	Secunderabad	21:15	21:31	2	
	2367	Guntakal	02:58	03:00	3	
		Bengaluru	07:55	-	3	

Sno	Train No.	Running Days				
		Tu	F	Su		
7	12265					
	Dist.					
	0	Stn. Name	Arr.	Dep.	Day	
	314	New Delhi	-	22:15	1	
	578	Ludhiana	02:45	02:55	2	
		Jammu Tawi	07:25	-	2	

Sno	Train No.	Running Days				
		Tu				
8	02283					
	Dist.					
	0	Stn. Name	Arr.	Dep.	Day	
	193	Ernakulam	-	23:25	1	
	419	Kozhikode	02:47	02:50	2	
	850	Mangalore	06:10	06:20	2	
	1184	Madgaon	12:50	13:00	2	
	1547	Ratnagiri	17:11	17:15	2	
	1958	Mumbai	22:40	22:45	2	
		Vadodara	05:13	05:31	3	

Chapter 1. Introduction

		2219	Ratlam	09:10	09:15	3
		2486	Kota	12:50	12:56	3
		2943	New Delhi	19:40	-	3
Sno	Train No.		Running Days			
9	12224		W Su			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Ernakulam	-	23:25	1
		193	Kozhikkode	02:47	02:50	2
		419	Mangalore	06:10	06:20	2
		850	Madgaon	12:50	13:00	2
		1184	Ratnagiri	17:11	17:15	2
		1599	Mumbai	23:56	-	2
Sno	Train No.		Running Days			
10	02284		Sa			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	New Delhi	-	21:36	1
		458	Kota	02:55	03:05	2
		724	Ratlam	07:05	07:10	2
		986	Vadodara	11:01	11:20	2
		1397	Mumbai	17:56	18:00	2
		1759	Ratnagiri	23:20	23:25	2
		2094	Madgaon	04:25	04:35	3
		2524	Mangalore	10:15	10:25	3
		2750	Kozhikkode	13:27	13:31	3
		2943	Ernakulam	18:20	-	3
Sno	Train No.		Running Days			
11	02286		M F			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	New Delhi	-	15:45	1
		403	Jhansi	20:22	20:27	1
		700	Bhopal	00:01	00:06	2
		1083	Nagpur	05:15	05:20	2
		1294	Balharshah	08:10	08:15	2
		1660	Secunderabad	14:00	-	2
Sno	Train No.		Running Days			
12	12264		M Th			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	New Delhi	-	10:56	1
		458	Kota	15:20	15:31	1
		724	Ratlam	18:36	18:40	1
		986	Vadodara	22:29	22:38	1
		1330	Mumbai	03:50	03:55	2
		1456	Lonavla	06:04	06:08	2
		1520	Pune	07:10	-	2
Sno	Train No.		Running Days			
13	12270		Tu Sa			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	New Delhi	-	15:45	1
		306	Gwalior	19:09	19:11	1

1.2 Sample datasets

41

		403	Jhansi	20:22	20:27	1
		700	Bhopal	23:59	00:04	2
		1083	Nagpur	05:15	05:20	2
		1294	Balharshah	08:10	08:15	2
		1745	Vijayawada	14:20	14:25	2
		2175	Chennai	20:40	—	2
Sno	Train No.		Running Days			
14	02245		Tu W F Sa Su			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Kolkata	—	10:50	1
		439	Bhubaneswar	16:50	17:00	1
		1219	Vijayawada	04:20	04:25	2
		1595	Renigunta	10:10	10:15	2
		1946	Bengaluru	16:00	—	2
Sno	Train No.		Running Days			
15	12222		Th Sa			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Kolkata	—	08:20	1
		250	Tatanagar	11:33	11:43	1
		720	Bilaspur	17:36	17:45	1
		1131	Nagpur	23:15	23:20	1
		1524	Bhusaval	03:42	03:47	2
		1708	Manmad	06:00	06:04	2
		1944	Daund	10:00	10:15	2
		2020	Pune	11:45	—	2
Sno	Train No.		Running Days			
16	12262		M Tu W F			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Kolkata	—	08:20	1
		250	Tatanagar	11:33	11:43	1
		720	Bilaspur	17:36	17:45	1
		1131	Nagpur	23:15	23:20	1
		1524	Bhusaval	03:42	03:47	2
		1831	Igatpuri	07:40	07:45	2
		1968	Mumbai	10:31	—	2
Sno	Train No.		Running Days			
17	12273		M F			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Kolkata	—	08:35	1
		200	Asansol	10:56	11:00	1
		311	Jasidih	12:25	12:27	1
		532	Patna	16:20	16:31	1
		744	Varanasi	20:13	20:23	1
		1090	Kanpur	00:40	00:45	2
		1529	New Delhi	06:25	—	2
Sno	Train No.		Running Days			
18	12228		F Su			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Indore	—	23:00	1

		80	Ujjain	00:15	00:20	2
		176	Ratlam	01:50	01:55	2
		438	Vadodara	05:20	05:30	2
		829	Mumbai	10:50	-	2
Sno	Train No.		Running Days			
19	12240		Tu Th			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Jaipur	-	19:11	1
		132	Sawai Madhopur	20:56	21:06	1
		506	Ratlam	01:50	01:55	2
		767	Vadodara	05:20	05:30	2
		1159	Mumbai	10:50	-	2
Sno	Train No.		Running Days			
20	12266		M W Sa			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Jammu Tawi	-	19:20	1
		265	Ludhiana	23:15	23:25	1
		578	New Delhi	04:20	-	2
Sno	Train No.		Running Days			
21	12219		W Sa			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Mumbai	-	23:06	1
		176	Pune	02:05	02:10	2
		422	Solapur	05:57	06:00	2
		773	Secunderabad	11:06	-	2
Sno	Train No.		Running Days			
22	12223		Tu Sa			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Mumbai	-	20:50	1
		415	Ratnagiri	02:50	02:55	2
		750	Madgaon	07:10	07:20	2
		1180	Mangalore	12:10	12:20	2
		1406	Kozhikode	15:23	15:25	2
		1599	Ernakulam	20:15	-	2
Sno	Train No.		Running Days			
23	12293		M F			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Mumbai	-	17:25	1
		121	Igatpuri	19:25	19:31	1
		429	Bhusaval	23:00	23:06	1
		981	Jabalpur	05:50	06:00	2
		1170	Satna	08:31	08:40	2
		1348	Prayagraj.	12:45	-	2
Sno	Train No.		Running Days			
24	12269		M F			
		Dist.	Stn. Name	Arr.	Dep.	Day
		0	Chennai	-	06:40	1
		431	Vijayawada	12:20	12:31	1

		880	Balbarshah	18:06	18:11	1
		1091	Nagpur	21:00	21:06	1
		1481	Bhopal	02:18	02:20	2
		1778	Jhansi	05:30	05:35	2
		1875	Gwalior	06:51	06:53	2
		2180	New Delhi	10:40	-	2
Sno	Train No.		Running Days			
25	12227		Th Sa			
	Dist.		Stn. Name	Arr.	Dep.	Day
	0		Mumbai	-	23:15	1
	392		Vadodara	03:52	04:05	2
	653		Ratlam	07:25	07:30	2
	750		Ujjain	08:50	08:56	2
	829		Indore	11:06	-	2
Sno	Train No.		Running Days			
26	12239		Tu Su			
	Dist.		Stn. Name	Arr.	Dep.	Day
	0		Mumbai	-	23:15	1
	392		Vadodara	03:52	04:05	2
	653		Ratlam	07:35	07:40	2
	1027		Sawai Madhopur	12:10	12:20	2
	1159		Jaipur	14:35	-	2
Sno	Train No.		Running Days			
27	12267		M Tu W Th F Sa Su			
	Dist.		Stn. Name	Arr.	Dep.	Day
	0		Mumbai	-	23:25	1
	491		Ahmedabad	05:55	06:15	2
	622		Surendranagar	08:24	08:34	2
	738		Rajkot	10:56	-	2
Sno	Train No.		Running Days			
28	22209		M F			
	Dist.		Stn. Name	Arr.	Dep.	Day
	0		Mumbai	-	23:15	1
	392		Vadodara	03:52	04:05	2
	653		Ratlam	07:35	07:40	2
	920		Kota	10:25	10:31	2
	1384		New Delhi	16:31	-	2
Sno	Train No.		Running Days			
29	12290		M Tu W Th F Sa Su			
	Dist.		Stn. Name	Arr.	Dep.	Day
	0		Nagpur	-	20:40	1
	393		Bhusaval	01:20	01:25	2
	701		Igatpuri	05:10	05:15	2
	837		Mumbai	08:06	-	2
Sno	Train No.		Running Days			
30	12274		Tu Sa			
	Dist.		Stn. Name	Arr.	Dep.	Day
	0		New Delhi	-	12:56	1

Chapter I. Indian Railways

44

440	Kanpur	17:48	17:54	1
786	Varanasi	22:11	22:20	1
998	Patna	01:10	01:20	2
1219	Jasidih	05:06	05:09	2
1330	Asansol	07:35	07:40	2
1529	Kolkata	10:40	-	2

Sno 31	Train No. 12282	Running Days		
		Th	Arr.	Dep.
		Stn. Name		
		New Delhi	-	12:56
		Kanpur	17:48	17:54
		Varanasi	21:50	22:00
		Adra	03:40	03:50
		Balasore	07:33	07:38
		Bhubaneswar	11:15	-

Sno 32	Train No. 22210	Running Days		
		Tu Sa	Arr.	Dep.
		Stn. Name		
		New Delhi	-	23:25
		Kota	04:15	04:25
		Ratlam	07:30	07:35
		Vadodara	11:10	11:20
		Mumbai	16:15	-

Sno 33	Train No. 02214	Running Days		
		Tu Th Sa	Arr.	Dep.
		Stn. Name		
		Patna	-	20:40
		Asansol	01:56	02:01
		Kolkata	05:40	-

Sno 34	Train No. 12294	Running Days		
		Tu Sa	Arr.	Dep.
		Stn. Name		
		Prayagraj.	-	19:20
		Satna	22:00	22:11
		Jabalpur	00:35	00:45
		Bhusaval	08:56	09:00
		Igatpuri	12:31	12:35
		Mumbai	14:56	-

Sno 35	Train No. 12221	Running Days		
		M Sa	Arr.	Dep.
		Stn. Name		
		Pune	-	15:15
		Daund	16:20	16:36
		Manmad	20:31	20:33
		Bhusaval	22:40	22:45
		Nagpur	04:10	04:15
		Bilaspur	09:56	10:06
		Tatanagar	15:56	16:06
		Kolkata	19:50	-

1.2 Sample datasets

45

Sno	Train No.	Running Days
36	12263	Tu F
	Dist.	Stn. Name
	0	Pune
	64	Lonavla
	191	Mumbai
	535	Vadodara
	797	Ratlam
	1063	Kota
	1520	New Delhi
Sno	Train No.	Running Days
37	12298	M Th Sa
	Dist.	Stn. Name
	0	Pune
	64	Lonavla
	191	Mumbai
	635	Ahmedabad
Sno	Train No.	Running Days
38	12268	M Tu W Th F Sa Su
	Dist.	Stn. Name
	0	Rajkot
	117	Surendranagar
	247	Ahmedabad
	738	Mumbai
Sno	Train No.	Running Days
39	02201	M W F
	Dist.	Stn. Name
	0	Kolkata
	136	Kharagpur
	459	Bhubaneswar
Sno	Train No.	Running Days
40	12259	M W Th Su
	Dist.	Stn. Name
	0	Kolkata
	266	Dhanbad
	667	Varanasi
	1014	Kanpur
	1453	New Delhi
	1628	Loharu
	1678	Sadulpur
	1736	Churu
	1779	Ratangarh
	1844	Sri Dungargarh
	1916	Bikaner
Sno	Train No.	Running Days
41	02285	Th Su
	Dist.	Stn. Name
	0	Secunderabad

367	Balharshah	18:06	18:11	1
578	Nagpur	21:00	21:06	1
967	Bhopal	02:15	02:20	2
1264	Jhansi	05:30	05:35	2
1667	New Delhi	10:40	-	2

Sno	Train No.	Running Days			Day		
		Tu	F	Stn. Name	Arr.	Dep.	
42	12220			Secunderabad	-	23:06	1
		0		Solapur	03:57	04:00	2
		334		Pune	07:50	07:55	2
		597		Mumbai	11:06	-	2
		773					

Sno	Train No.	Running Days			Day			
		M	W	Sa		Stn. Name	Arr.	Dep.
43	22204				1	Secunderabad	-	20:15
		0			2	Vijayawada	01:05	01:15
		350			2	Visakhapatnam	06:30	-
		699						

Sno	Train No.	Running Days			Day			
		M	W	F		Stn. Name	Arr.	Dep.
44	02213				1	Kolkata	-	22:00
		0			2	Asansol	01:08	01:12
		215			2	Patna	06:40	-
		546						

Sno	Train No.	Running Days			Day			
		Tu	Th	Su		Stn. Name	Arr.	Dep.
45	22203				1	Visakhapatnam	-	19:45
		0			2	Vijayawada	00:40	00:55
		350			2	Secunderabad	06:15	-
		699						

Sno	Train No.	Running Days			Day				
		M	Tu	Th	F	Su	Stn. Name	Arr.	Dep.
46	02246					1	Bengaluru	-	11:00
		0					Renigunta	16:50	17:00
		352					Vijayawada	22:20	22:31
		728					Vizianagram	04:30	04:40
		1126					Bhubaneswar	09:50	10:00
		1507					Kolkata	16:45	-
		1946							

Sno	Train No.	Running Days			Day			
				Sa		Stn. Name	Arr.	Dep.
47	12213				1	Bengaluru	-	23:40
		0			2	Guntakal	03:45	03:50
		276			2	Secunderabad	08:56	09:10
		688			2	Balharshah	13:31	13:35
		1054			2	Bhopal	21:20	21:25
		1655			2	Jhansi	01:15	01:20
		1952			3			

1.2 Sample datasets

47

	2374	New Delhi	07:00	-	3
Sno	Train No.	Running Days			
48	12297	M Th Su			
	Dist.	Stn. Name	Arr.	Dep.	Day
	0	Ahmedabad	-	22:25	1
	444	Mumbai	03:50	03:55	2
	567	Lonavla	06:05	06:07	2
	631	Pune	07:10	-	2

Figure 1.9: Train dataset

Lecture 2 : Concept of variables, iterators and filtering

1. Variable

- A **variable** is like a container that stores a value which can change during a process.
- In CT, variables help us **store information temporarily** and reuse it.

Example:

If you want to store a student's age:

```
age = 20
```

2. Iteration

- **Iteration** means repeating a process (looping).
- Instead of doing the same work again and again manually, iteration helps automate repetition.

Example:

Print numbers from 1 to 5:

```
for i in range(1, 6):  
    print(i)
```

3. Filtering

- **Filtering** means selecting only the elements that match a given condition.
- It reduces unnecessary data and focuses on what's relevant.

Example:

From a list of numbers, select only the even ones:

```
numbers = [1, 2, 3, 4, 5, 6]

evens = [n for n in numbers if n % 2 == 0]

print(evens)
```

Summary in Simple Words:

- Variable → stores information (like a box with a label).
- Iteration → repeating steps (like doing the same action in a loop).
- Filtering → picking only what you need (like choosing ripe fruits from a basket).

In lecture Prof Go through each and Every card

Variable and Iteration

Ex uses by prof

Counting cards

Avg in Math Marks

Based on Filtering Condition

Pronoun , Verb , And Other Part of Speech

In tutorial instructor also taught same by All food and no food Example and Avg word length

* (Dataset) Ex() → In Any Order / Randomly
and filtering. (Syntactic)

* Iteration :- When we go through the sequence of Objects that we have and for each object we keep a count (we do something). Ex(Counting).

Iterator itself has a initialization in Ex case count equal to 0 is the initialization.

* Variables :- An entity whose value keep changing as the computation goes on.

$$\text{Initialization} - \text{Count} = 0, \text{Lap} + \text{Count} = 5 - \text{Variables}$$

Ex in video - Counting Cards
Average in math marks. Sum 271
Count = 50] each

* Filtering :- To take out/ select specific type of data from entire dataset.

We take out data and pass it through filter and anything which is caught by filter stays behind and all the thing pass through. (Selective Iteration)

Note :- ① When moving a card from one file to other for counting the No of cards, the order in which the cards are processed is not important.

② We can do Avg Math marks Example in single iteration

Tutorial for ?

Going Through each Cards

Lecture 1.3 Iterations using combination of filtering conditions

- Iteration is Given condition with 'and' No matter in what order of condition you apply. # for in loop
if from class
 - Note - Best Way for short range of whole class, $\frac{\text{bogsum} + \text{gibsum}}{\text{bogcount} + \text{gibcount}}$ ① 1st half $\frac{1}{2}$ 2nd half
 - Declare variable, initialise = 0 ② Whole batch includes
Bog / Gib
- Note - The Pattern of doing something repetitively is called an ITERATOR ###
- * Description of Iterators.
- Step① Initialization Step : Arrange all the cards in an 'Unseen' pile
- Step② Continue or Exit : If there are no more cards in the 'Unseen' pile, we exit otherwise we continue.
- Step③ Repeat Step : Pick an element from the "Unseen" pile, do whatever we want to with this element and then move it to another "seen" pile.
- Step④ Go Back to Step 2.
- (Tutorial - Large Adjective ($S > M$) / 1st half / 2nd half, Lib)

Lecture 1.4 Introduction to Flowcharts

Introduction to Flowcharts

Flowcharts: Some commonly used symbols



Process or
Activity



Flowline or
Arrow



Decision



Terminal

1.

Flowcharts: Some commonly used symbols



Process or
Activity



Flowline or
Arrow



Decision



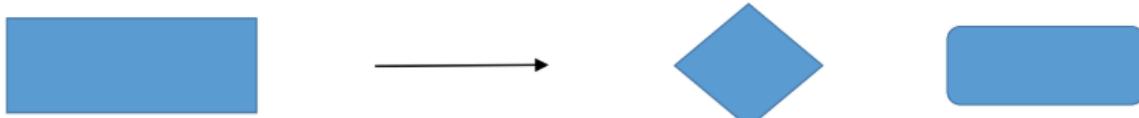
Terminal



Set of operations that
change the value of
data (variables)

2.

Flowcharts: Some commonly used symbols



Process or
Activity

Flowline or
Arrow

Decision

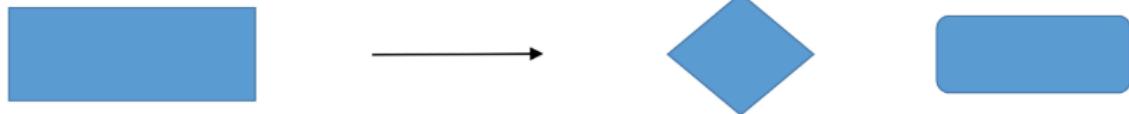
Terminal



Shows the order
of execution of
the program steps

3.

Flowcharts: Some commonly used symbols



Process or
Activity

Flowline or
Arrow

Decision

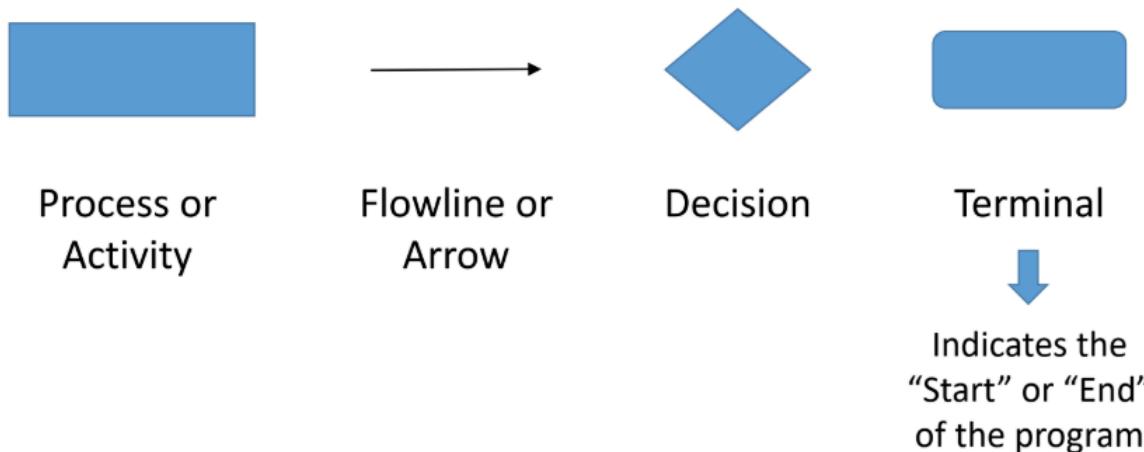
Terminal



Determines which
path the program
will take

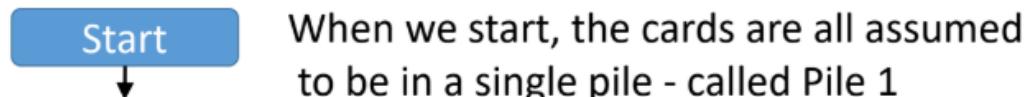
4.

Flowcharts: Some commonly used symbols

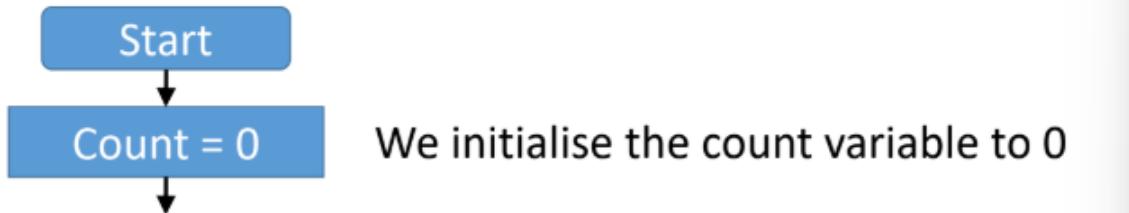


Ex 1 : Flowchart for counting cards

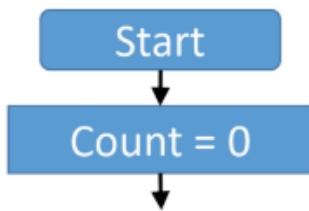
1.



2.



3.

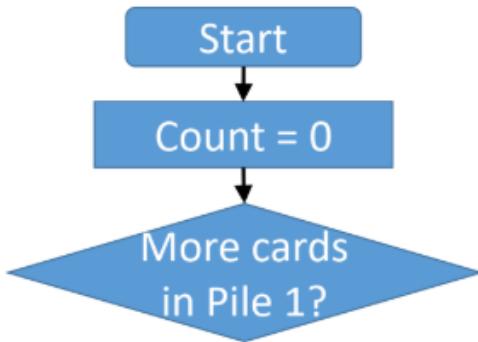


This is where the iterator starts.
We need to repeat the following steps:

- Pick a card from Pile 1
- Move it into a different pile (say Pile 2)
- Increment the value of count

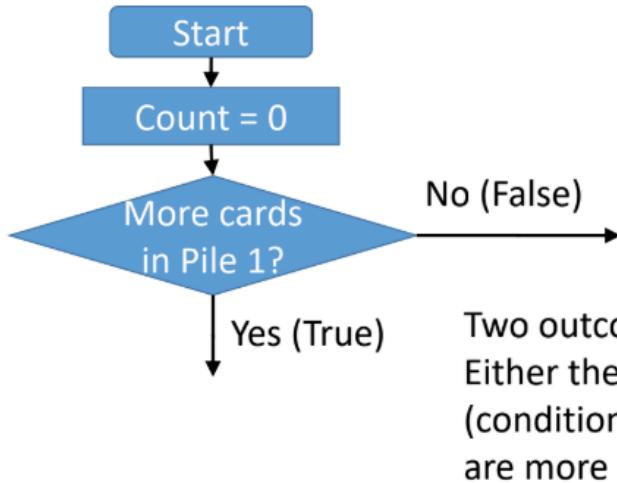
The iteration stops when Pile 1 is empty

4.



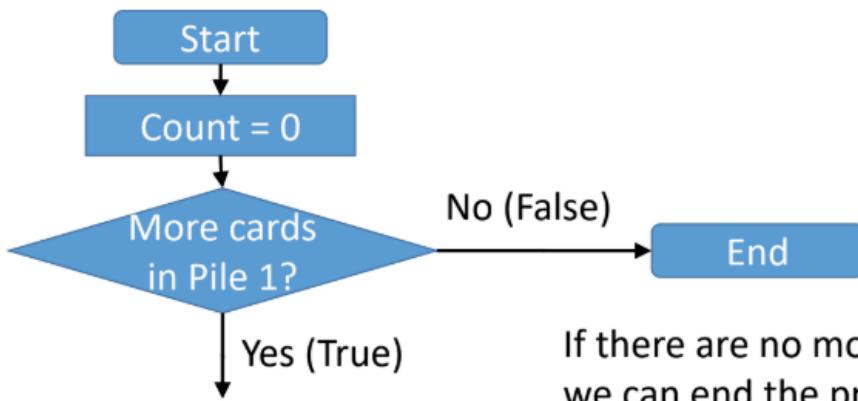
We need to pick a card from Pile 1
But we can do that only if there
are more cards in Pile 1 to pick !

5.



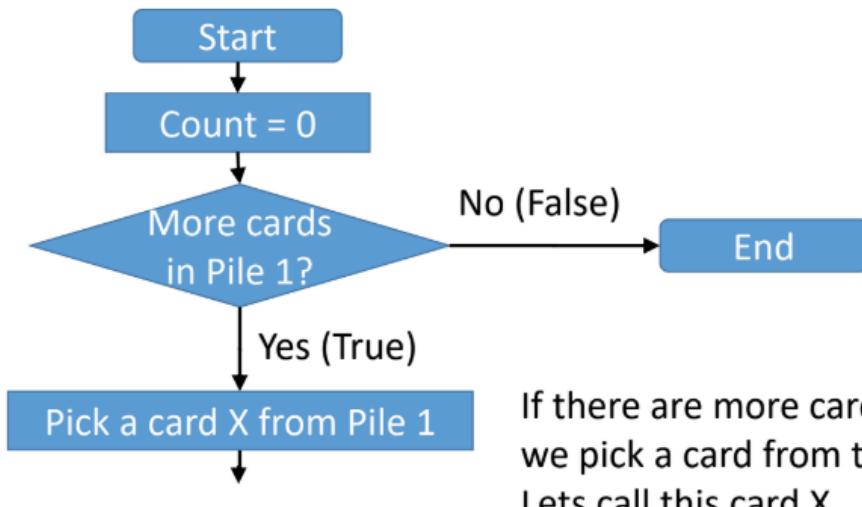
Two outcomes are possible here:
Either there are no more cards
(condition = False) or there
are more cards (condition = True)

6.



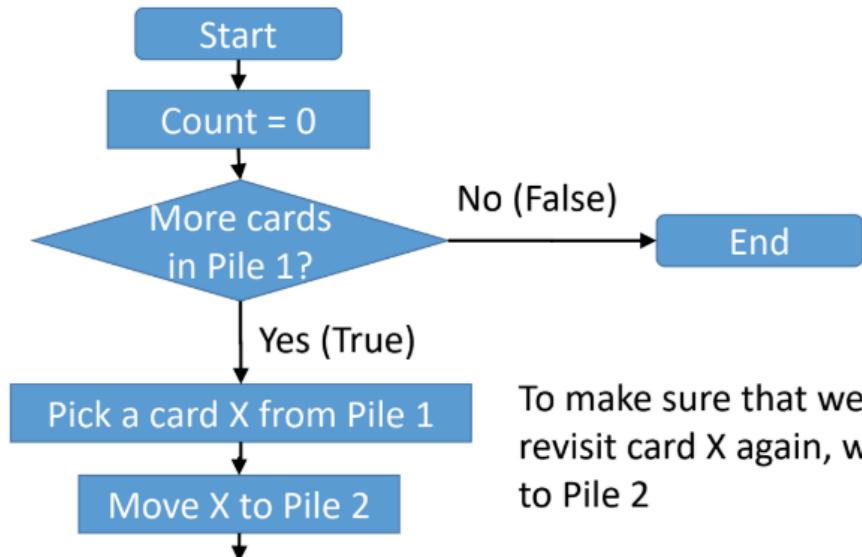
If there are no more cards
we can end the program.
The variable count carries the
required value.

7.



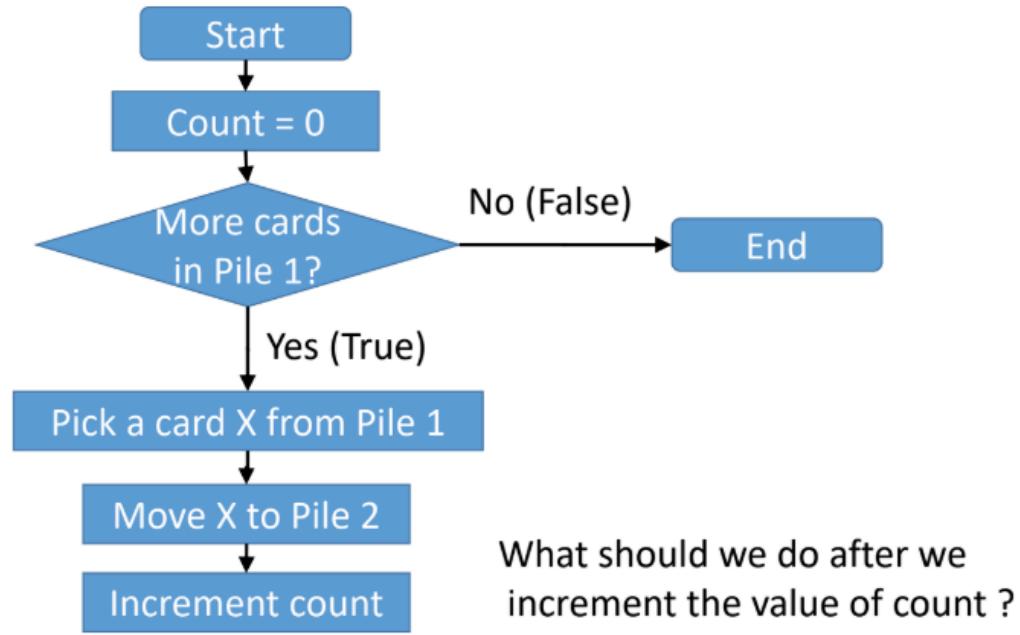
If there are more cards in Pile 1
we pick a card from the pile.
Lets call this card X

8.

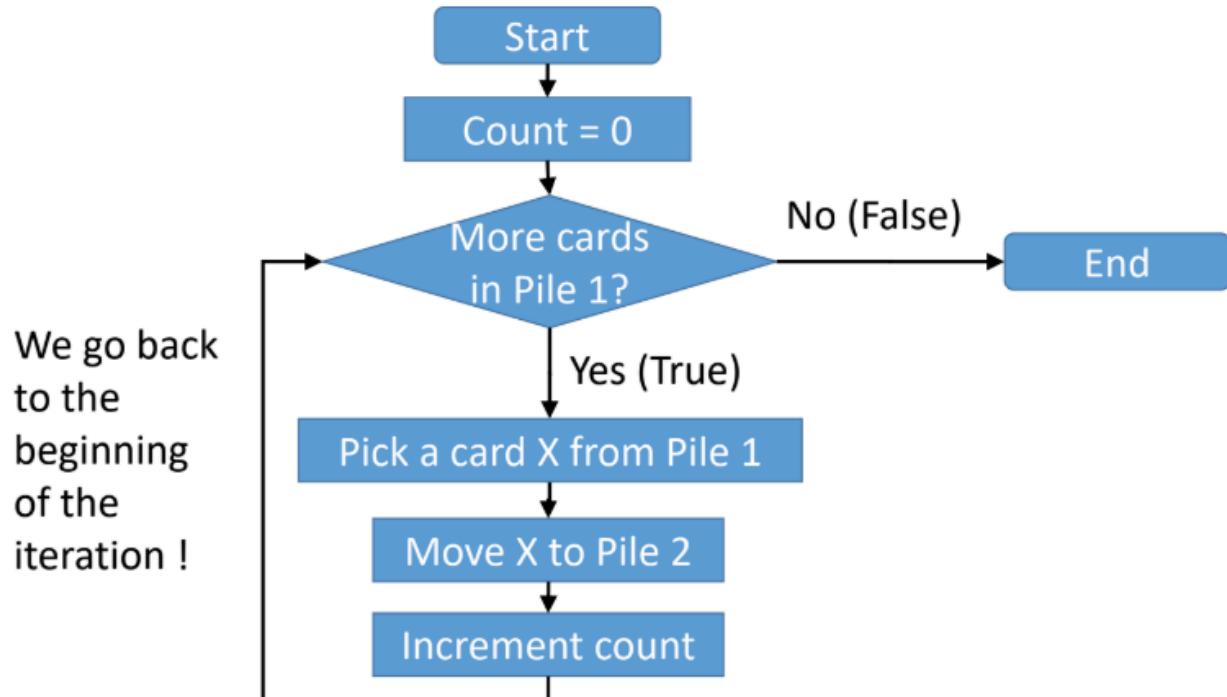


To make sure that we don't
revisit card X again, we move it
to Pile 2

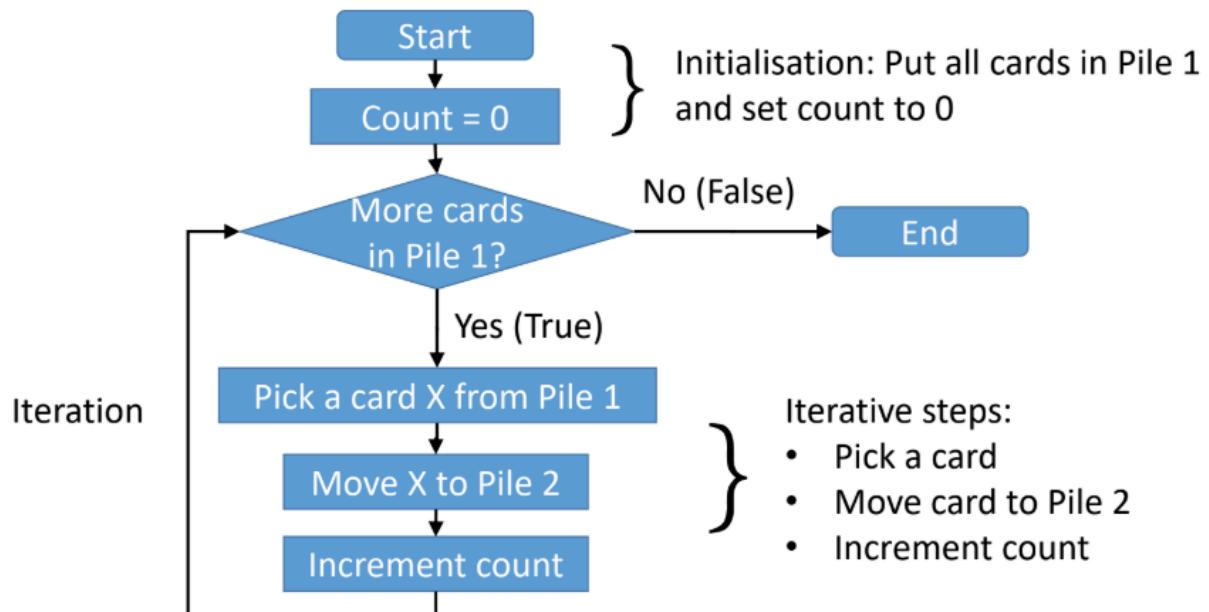
9.



10.

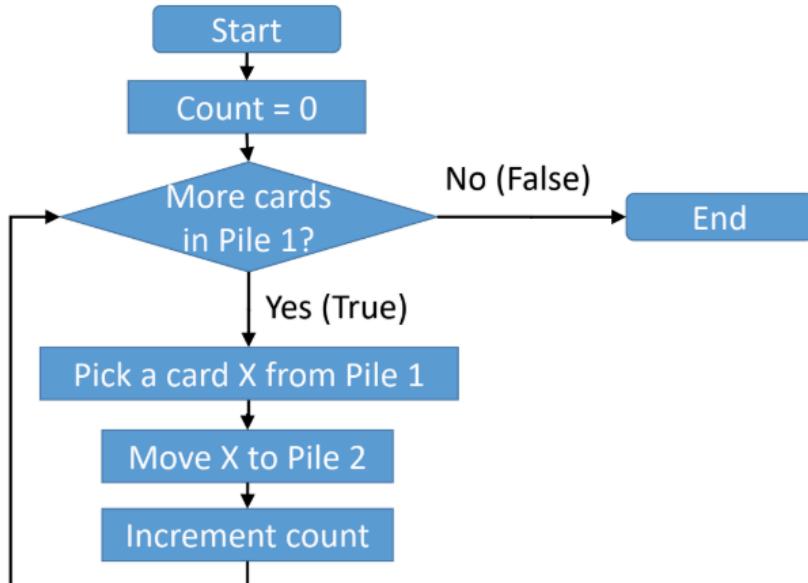


Summary :

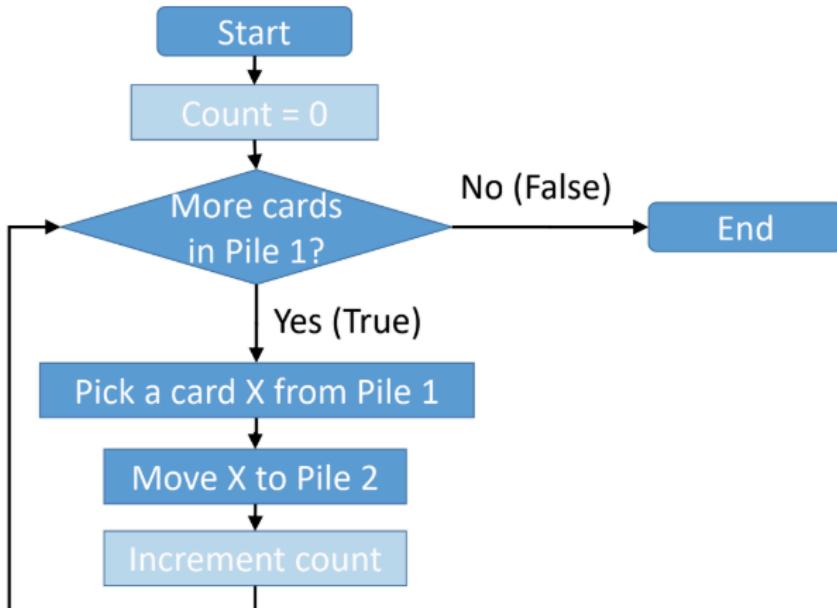


Ex 2 : Flowchart for Sum of Maths

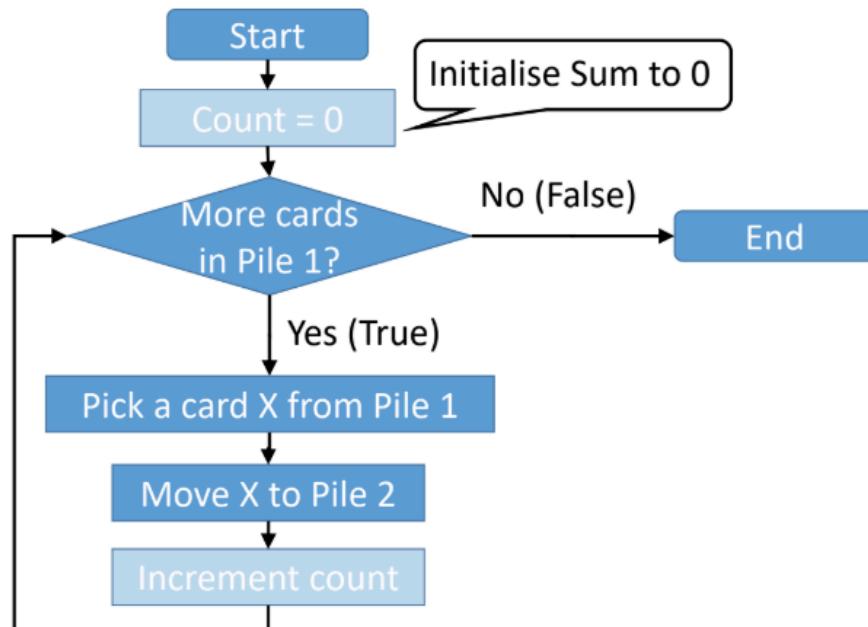
Flowchart for counting cards



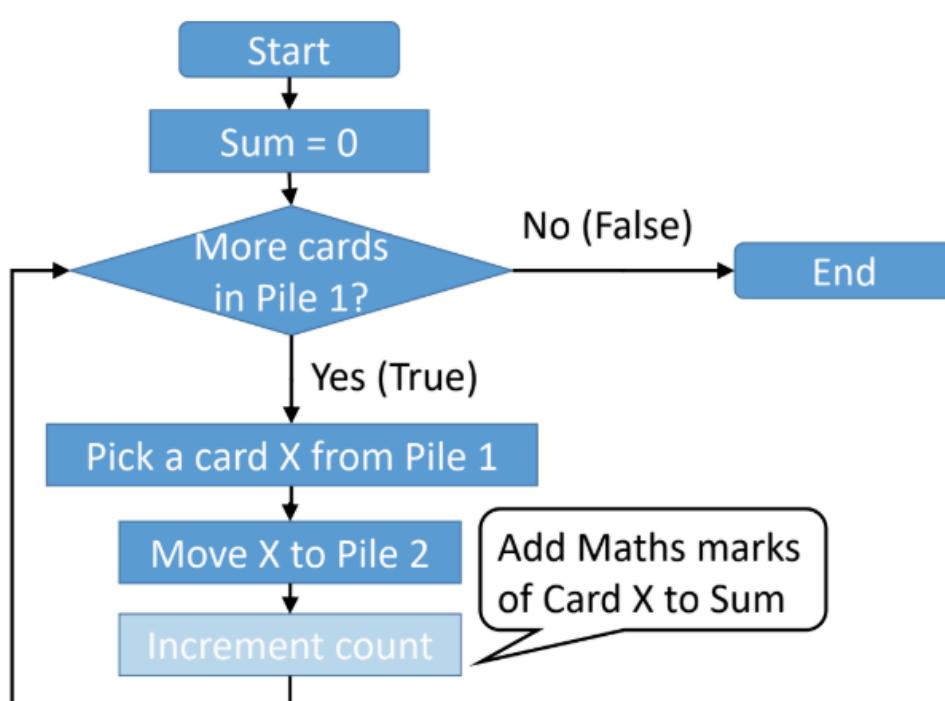
How do we modify this to do Sum of Maths marks?



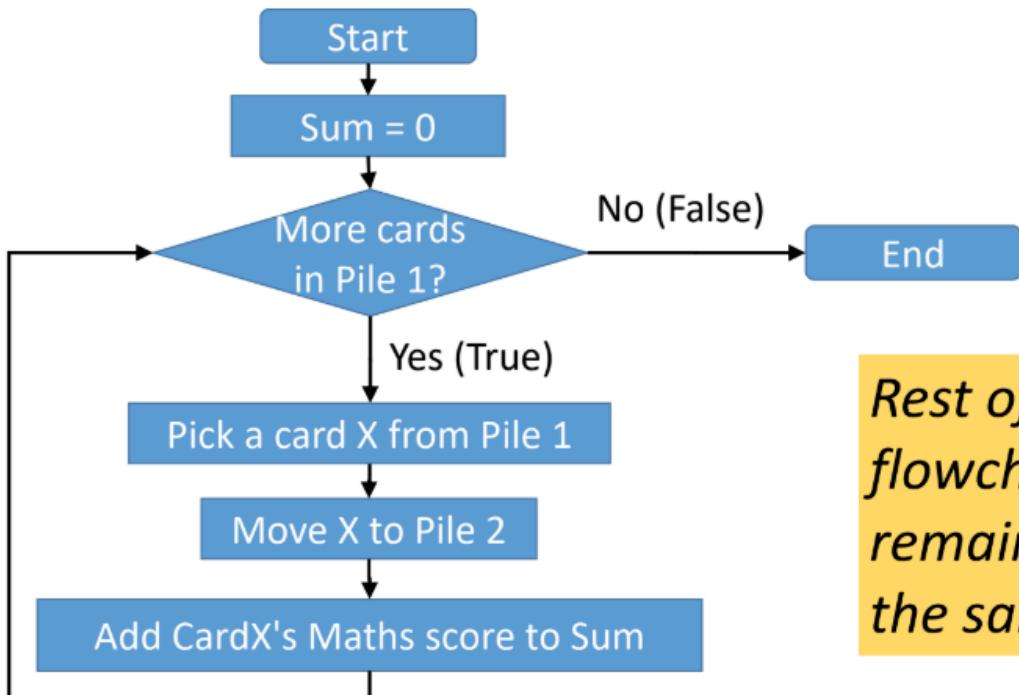
1.



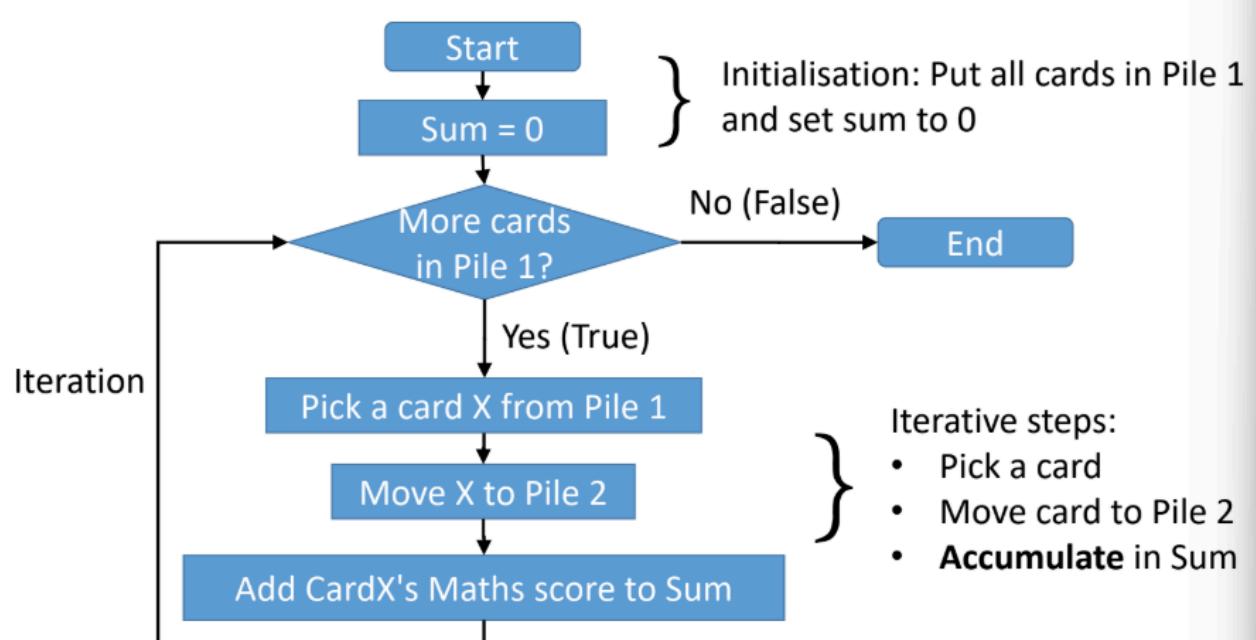
2.



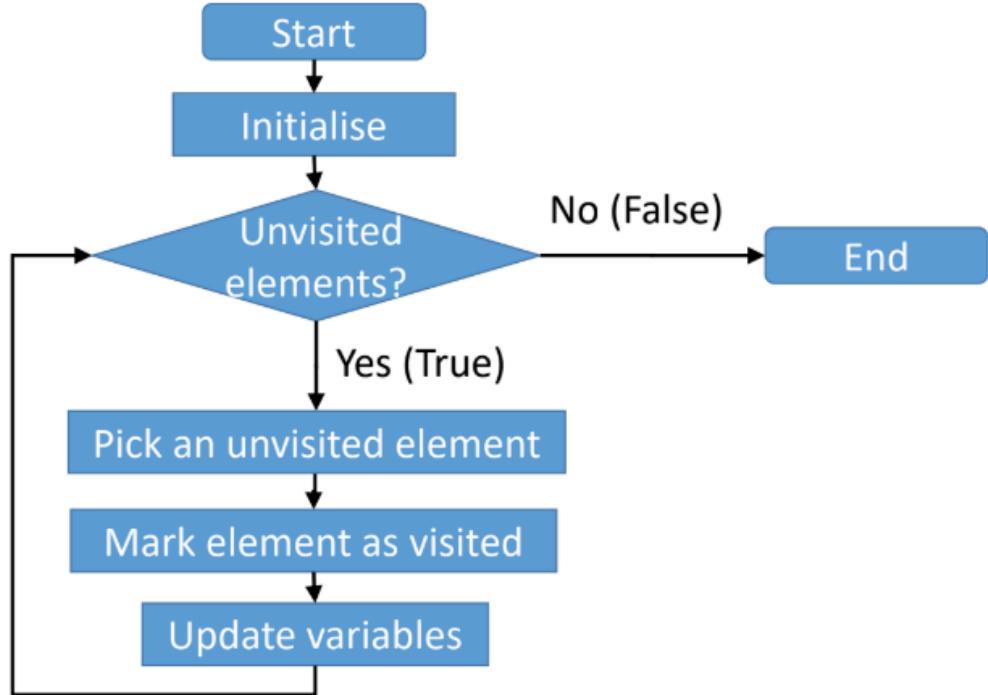
3.



Summary :



Summary: Generic flowchart for iteration



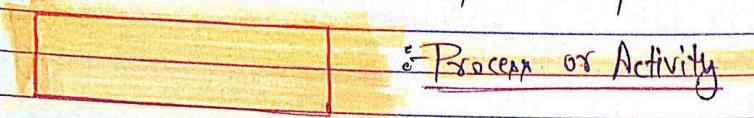
* Flowcharts :- (Definition)

These are basically diagrammatic representations of the sequence of steps that we are going to use, or algorithms that we are going to use.

The step-wise description of the program is visualised nicely using a diagram called flowchart.

* Flowcharts :- Some commonly used symbols :-

* ①



- Process or Activity

↓
Set of operation that change
the value of data (Variables). [In book [Any activity that needs
to be performed.]]

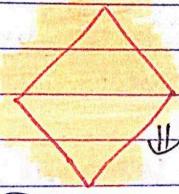
* ②



- Flowline or Arrow

↓
Show the order of execution of
the program steps. [In book [Shows program from one step
to other step].]

* ③



- Decision

↓
Determines which path the program
will take.

[In book -

Check some condition and branch
Based on true / false.

* ④



- Terminal

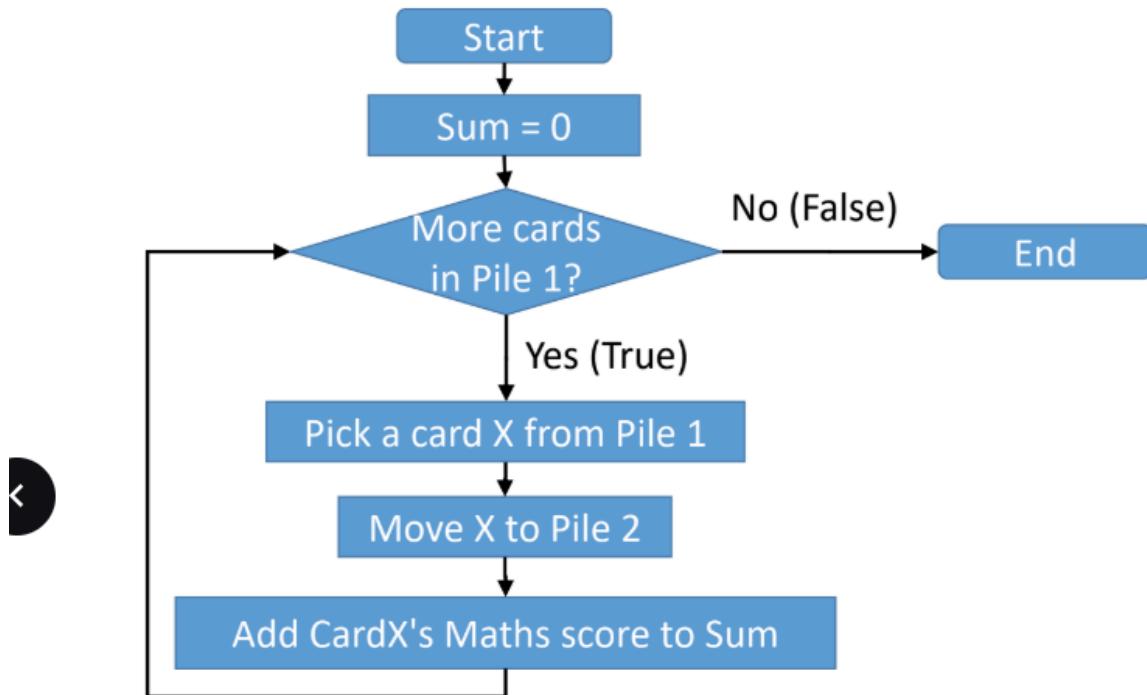
↓
Indicates the 'start' or 'end'
of the Program.

[The start or end of the
flowchart.

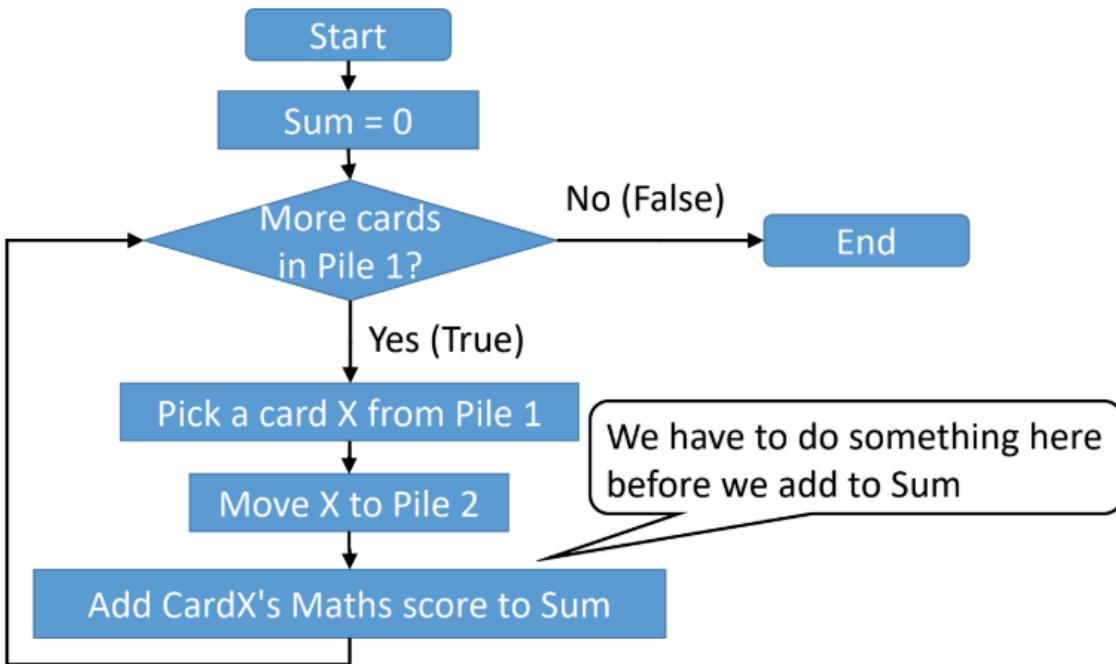
Lecture 1.5 Flowchart for Sum with Filtering

Flowchart for Sum with Filtering

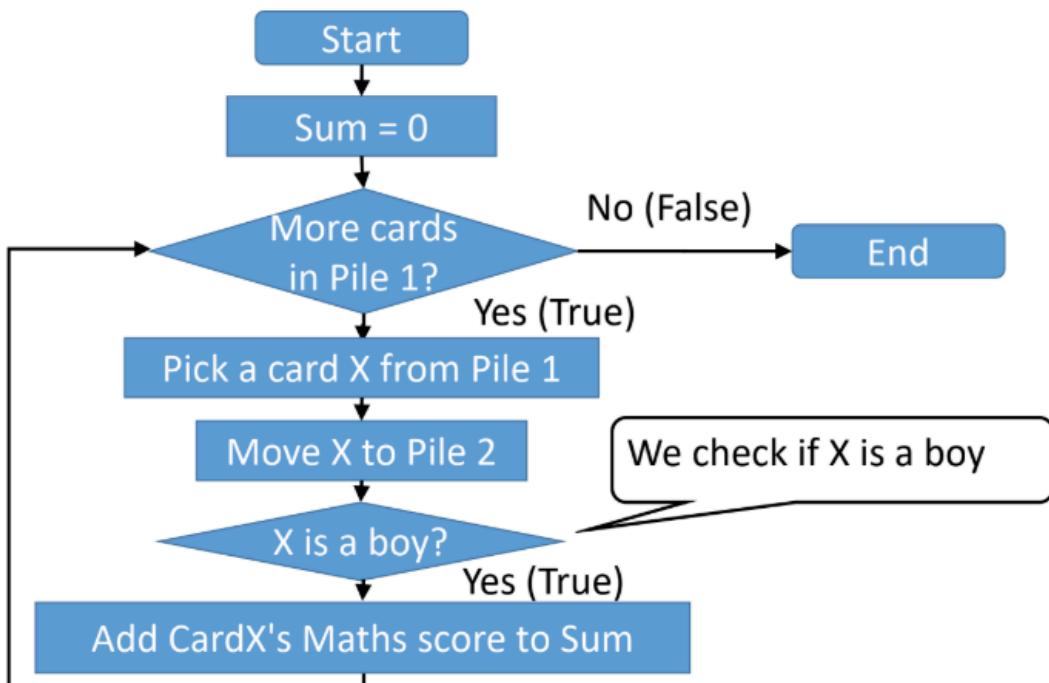
How do we modify this to do sum of Boys Maths marks?



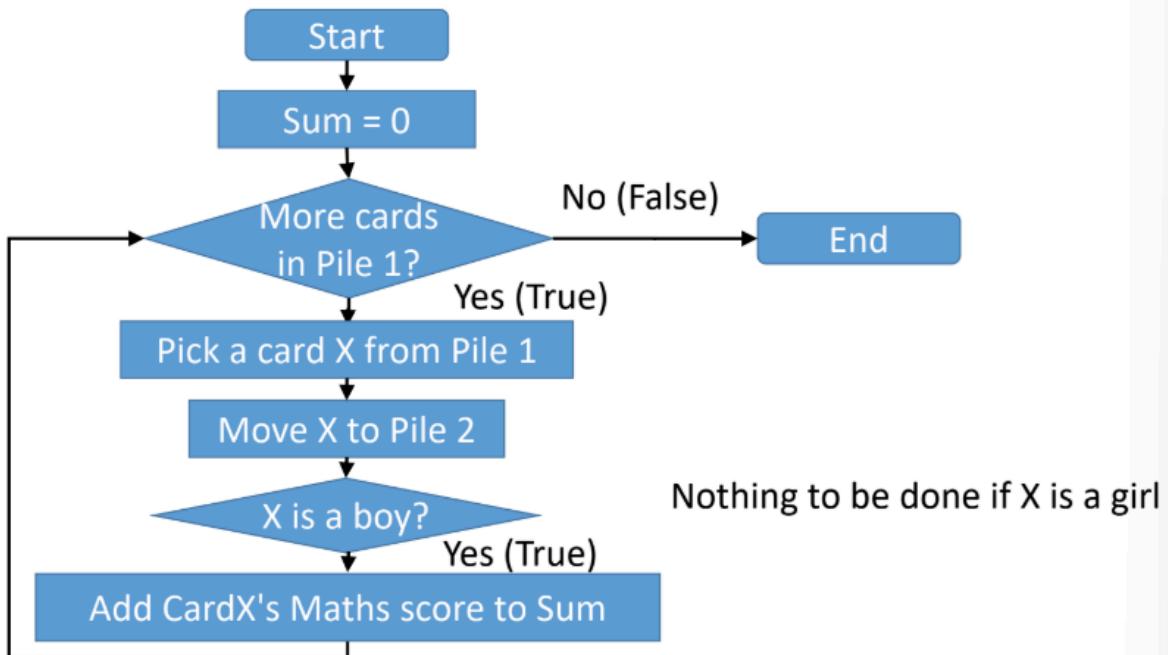
1.



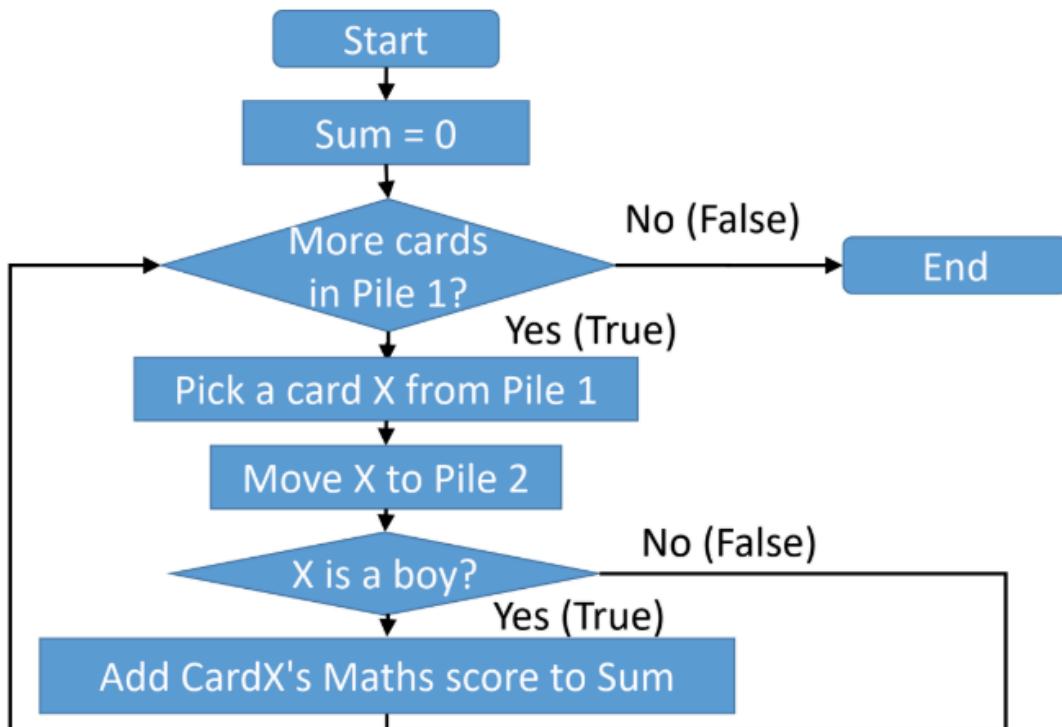
2.



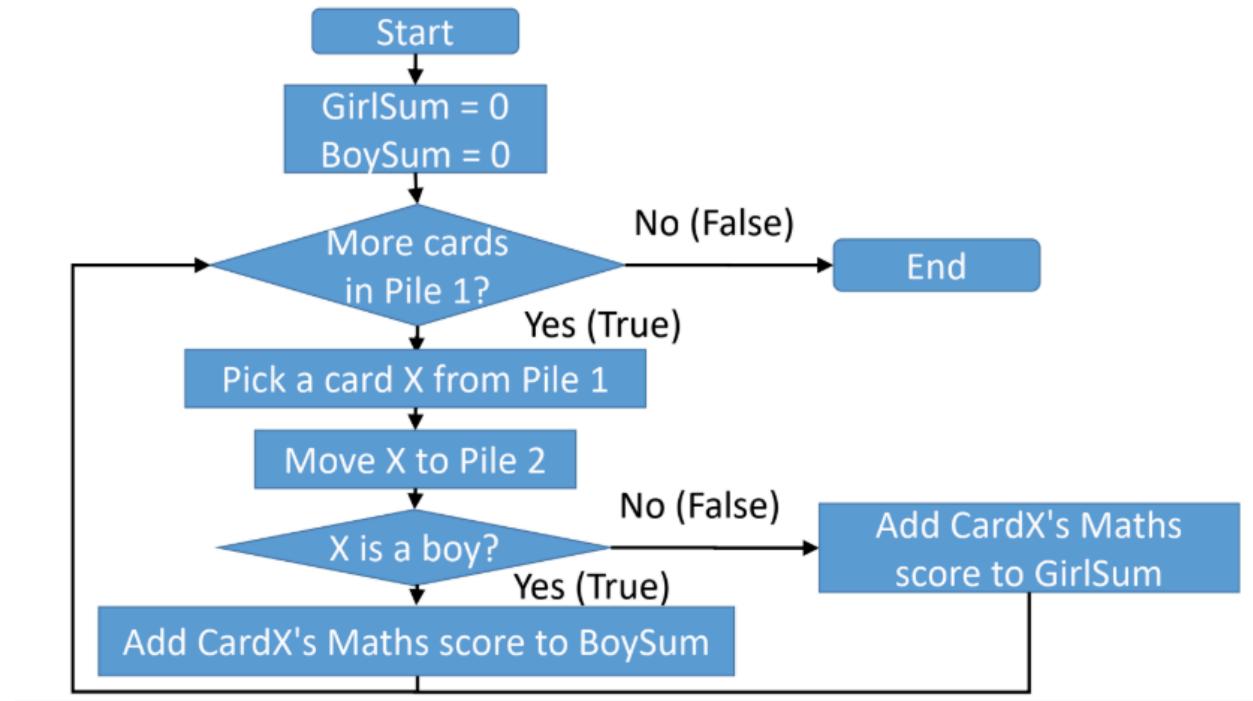
3.



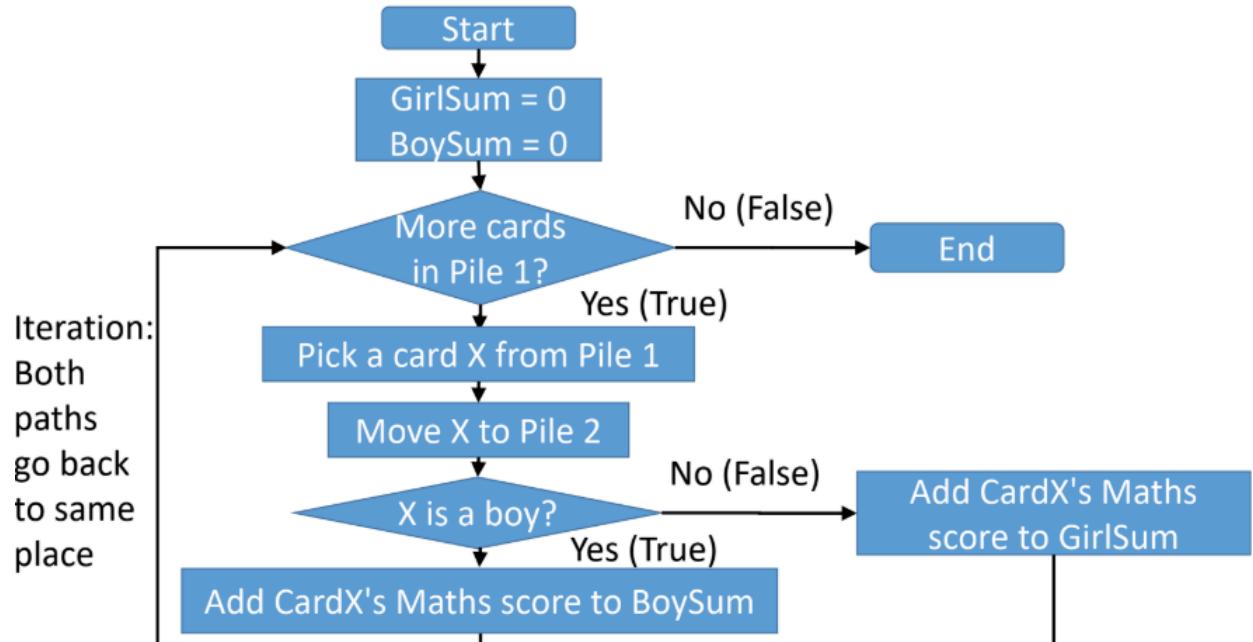
4.



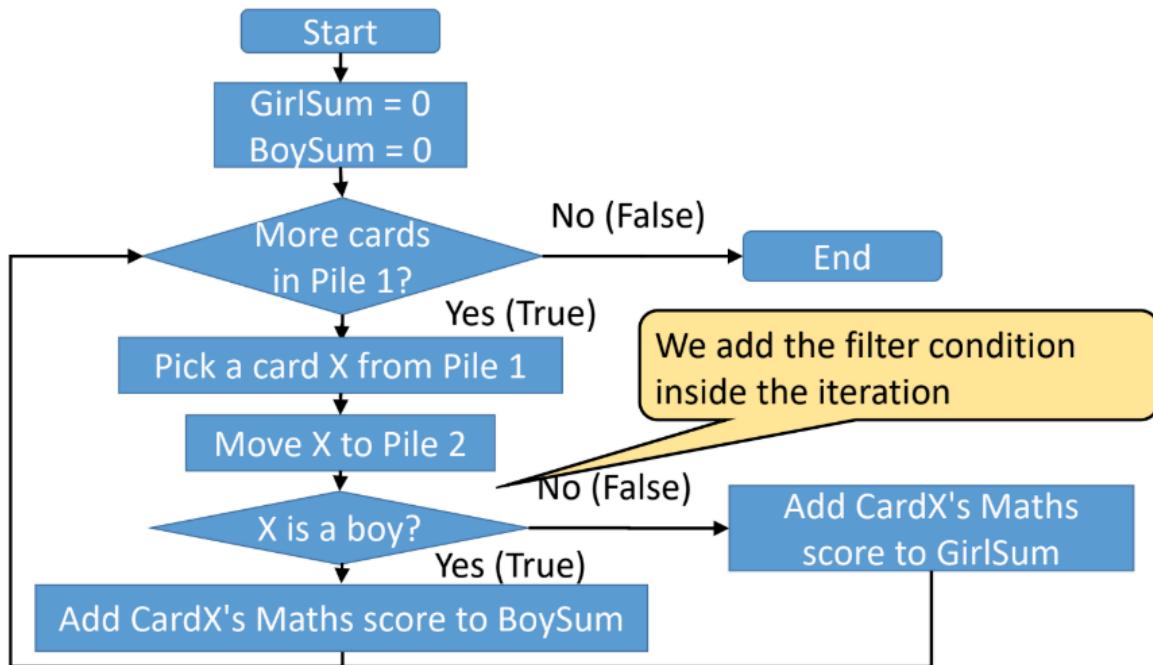
Sum of both Boys and Girls Math marks



2.



Summary : Sum with Filtering



Lecture 1.6 Sanity of data

Sanity of data: what we observed

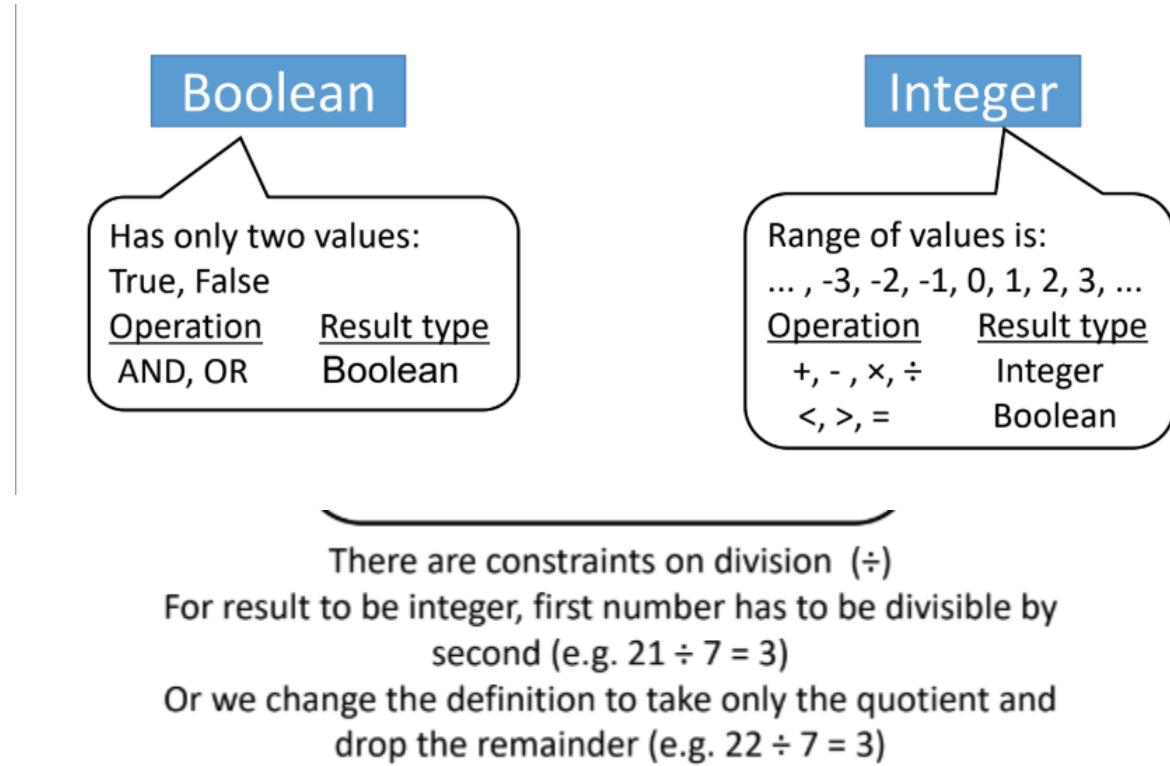
- We organised our data set into cards, each storing one data item
- Each card had a number of elements, e.g.:
 - numbers (e.g. marks)
 - sequence of characters (e.g. name, bill item, word etc)
- We observed that there were restrictions on the values each element can take:
 - for example marks has to lie between 0 and 100
 - name cannot have funny characters
- Constraints on the kinds of operations that can be performed:
 - addition of marks is possible
 - but multiplication of marks does not make sense !
 - compare one name with another to generate a **boolean type** (True or False)
 - but cannot add a name with another !

This leads us to the concept of a **Data Type** ...

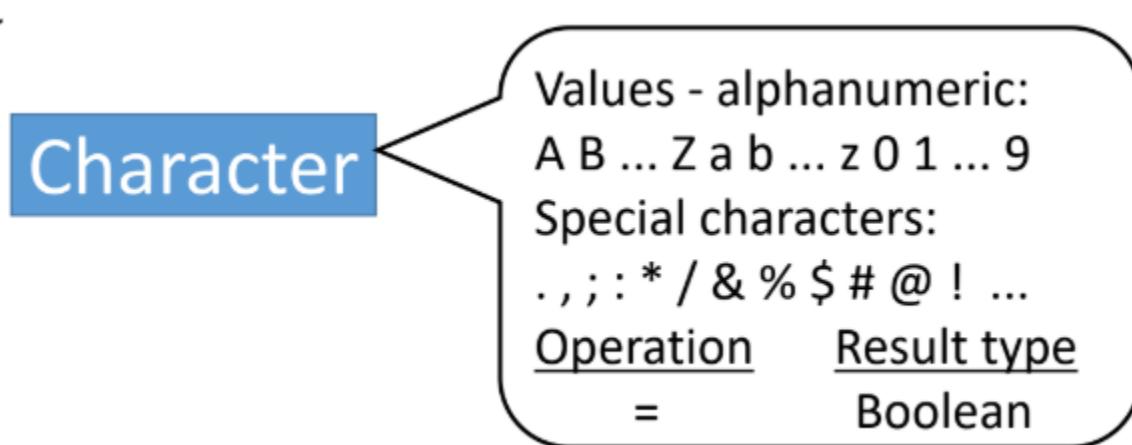
- By associating a **Data Type** (or simply **Type**) with a data element, we can tell the computer (or another person) how we intend to use a data element:
 - What are the values (or range of values) that the element can take ?
 - What are the operations that can be performed on the data element ?
- When we specify that a variable is of a specific type, we are describing the constraints placed on that variable in terms of the values it can store, and the operations that are permitted on it

Lecture 1.7 Introduction to Datatypes

Basic Data Types



Character Data Type



* Summary :-

* Boolean - Has 2 values (True and False)
Operation - And, Or, Not
Result type - Boolean

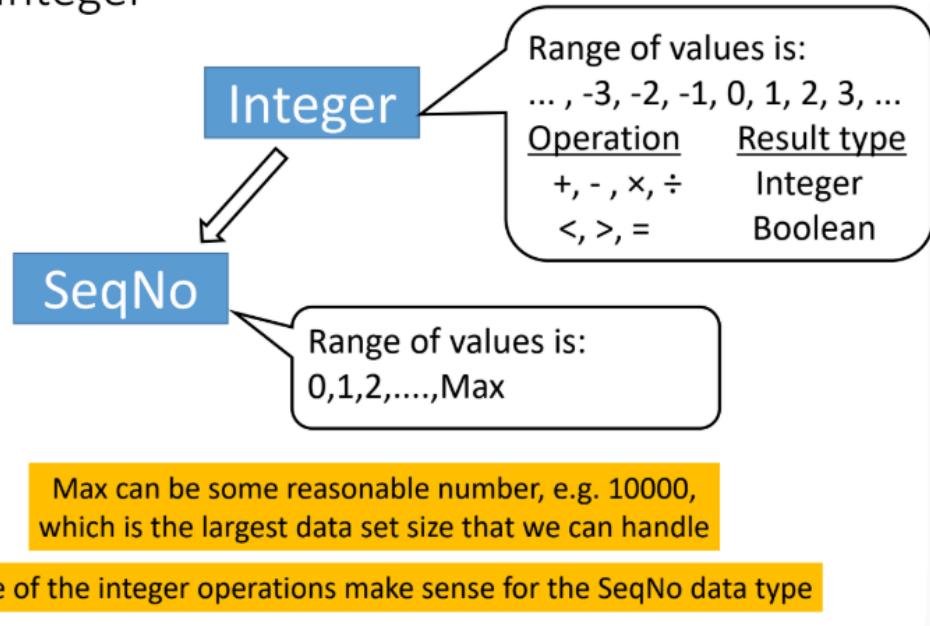
* Integer - Range of Values is -3, -2, -1, -0, 1, 2, ...
Operation : +, -, ×, ÷ ; <, >, =
Result type : Integer ; Boolean

* Character -
Values - alphanumeric.
A, B, C, ..., Z, a, b, ..., o, l, ..., g
Special Character - ., ;, :, *, /, %, #, \$, @, !
Operation =
Result type = Boolean

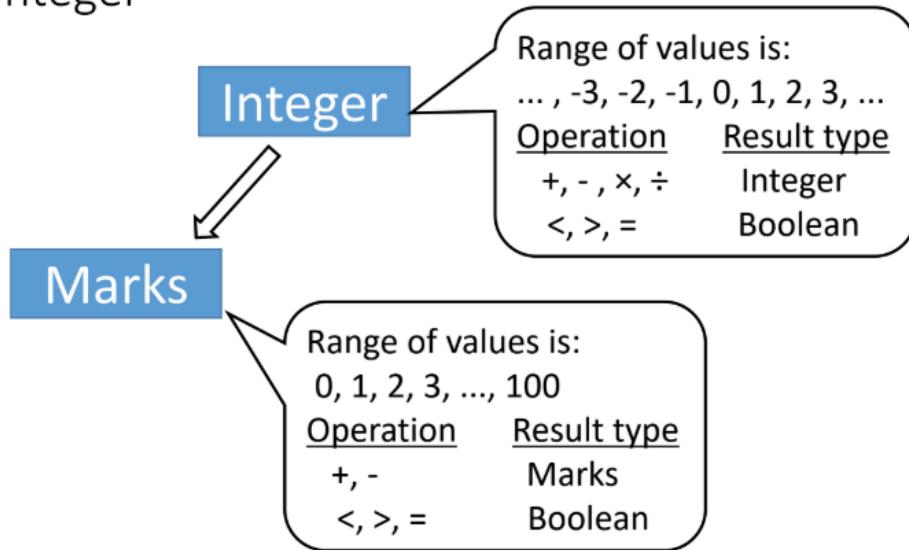
Lecture 1.8 Subtypes of basic datatypes

Subtypes of Integer

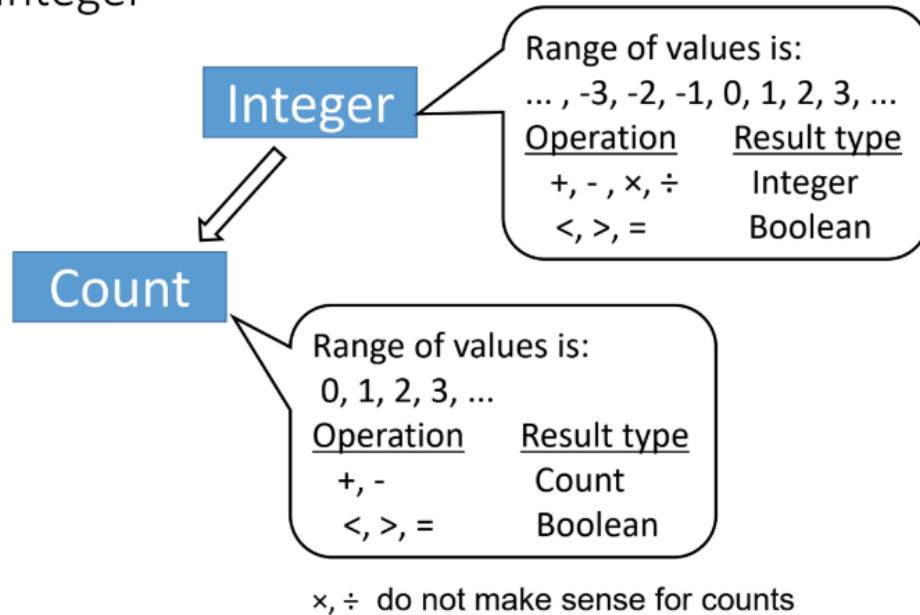
Subtypes of Integer



Subtypes of Integer

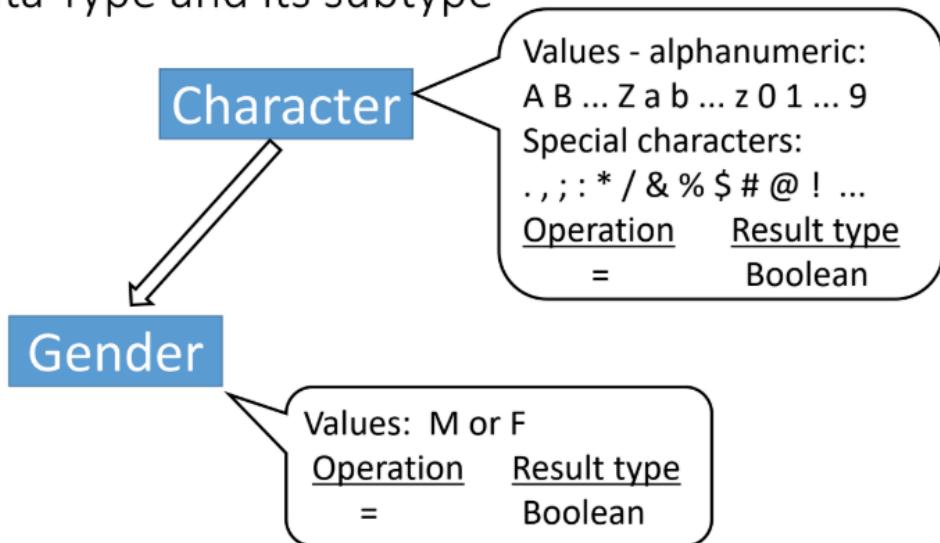


Subtypes of Integer



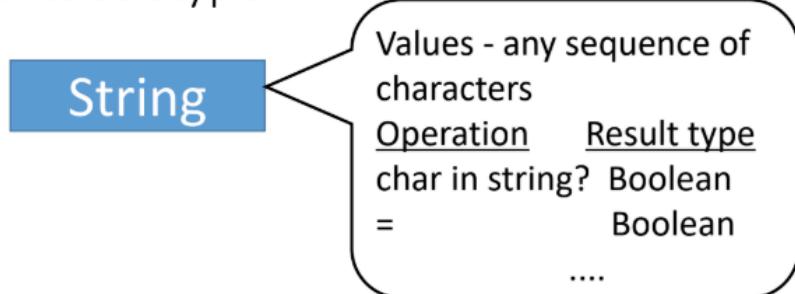
Character Data Type and Its subtype

Character Data Type and its subtype

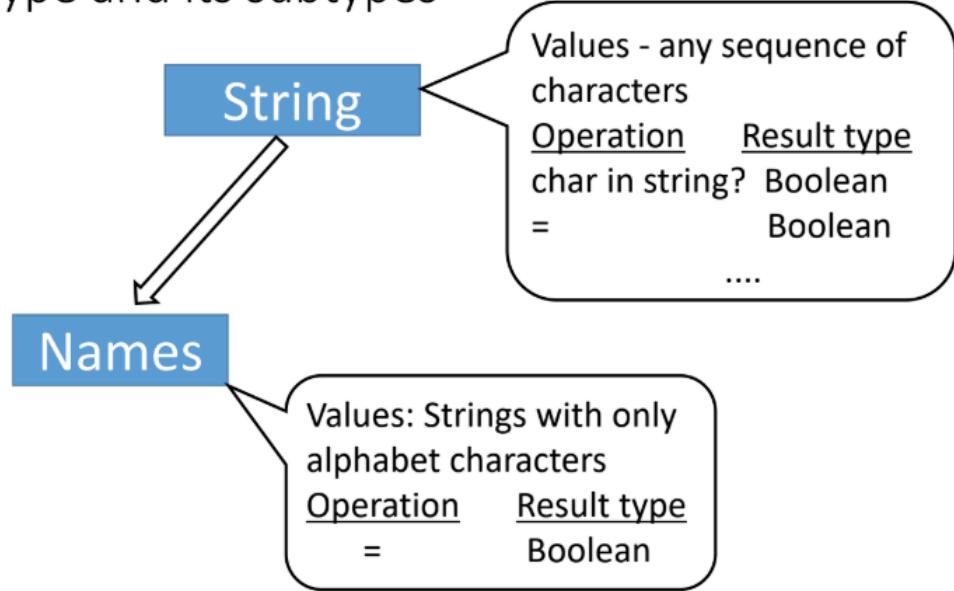


String Data Type and its Subtypes

String Data Type and its subtypes

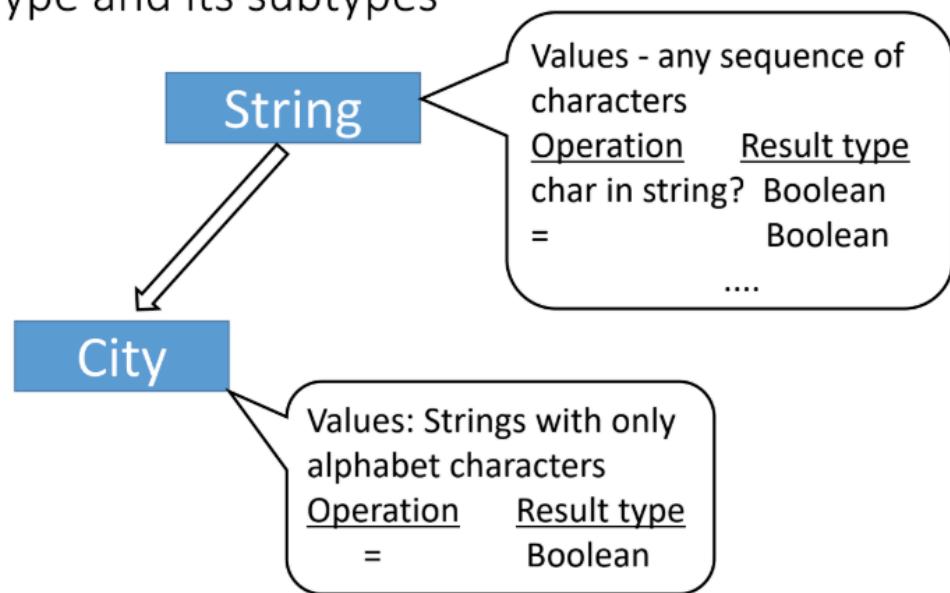


String Data Type and its subtypes



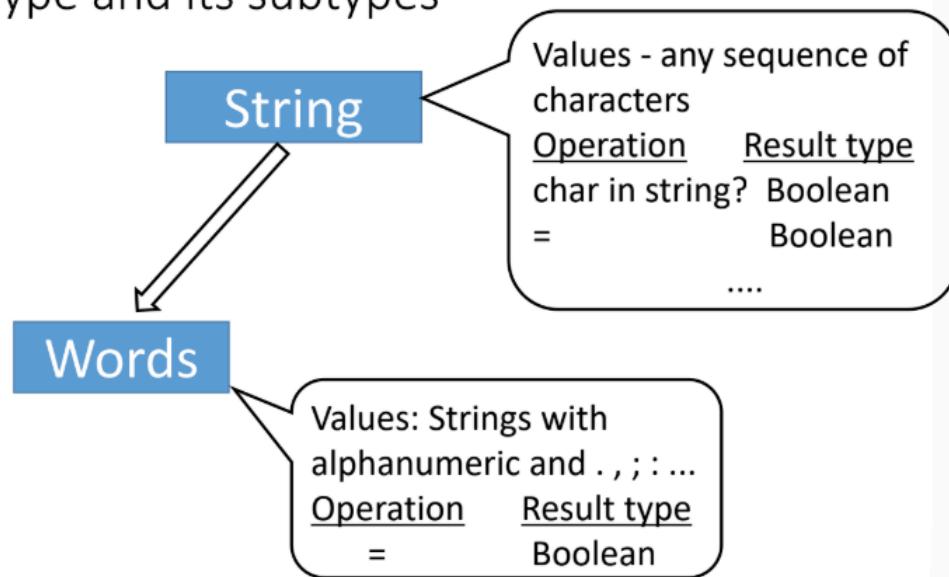
Note - Change from video: String made more restrictive and operation added

String Data Type and its subtypes



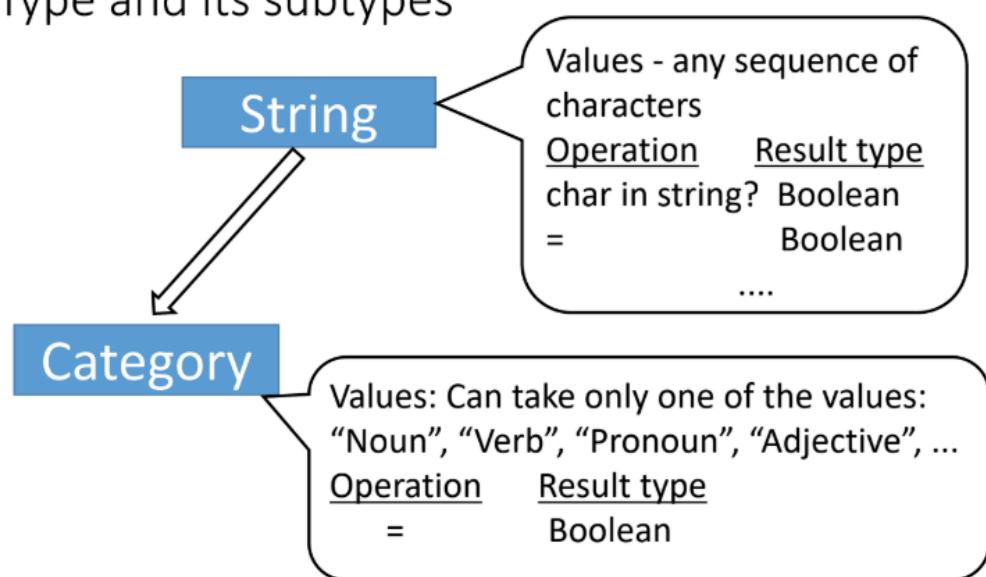
Note - Change from video: String made more restrictive and operation added

String Data Type and its subtypes



Note - Change from video: operation added

String Data Type and its subtypes

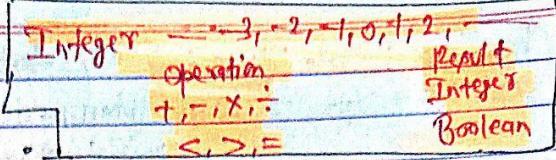


Note - Change from video: operation added

12/06/2024

L.1.8. Subtypes of Basic Datatype

* Subtypes :-



a) Subtypes of Integer :-

There are a number of places in our datasets where we have used integer with restriction. These are listed below :

(a) Sed No :- Should have only ~~few~~ Value (-ve and 0 are disallowed) and a max value for the upper limit of the range of data values.

* Range of Value = 0, 1, 2, ..., Max

* Operation = Boolean Operation

(b) Marks :- should have starting from 0.

Range : 0, 1, 2, 3, ..., 100

Operation = +, - | <, >, =

Result type = Marks | Boolean

(c) Count :- allows 0 or any positive integer value upto a sufficiently high limit max.

Range = 0, 1, 2, 3

Operation = +, - | <, >, =

Result type = Count | Boolean

(d) Hour : is a Subtype of ~~Interval~~ Integer with range 0 to 23. Addition, Subtraction, Comparison should be allowed.

(e) Min :- Range 0 to 59

(7) Day Interval :- Range 0 to 5.
Addition, Subtraction, Comparison allowed.

(8) Coord Range :- is a subtype of integers with range 0 to some maximum value (say 1000). Addition, Subtraction and Comparison should be allowed.

Summary table :-

Subtype	Values allowed	Operation allowed
① Seq No	0 — Max	=
② Marks	0 — 100	+, -, =, <, >
③ Count	0 — Max	+, -, =, <, >
④ Hours	0 — 23	"
⑤ Mins	0 — 59	"
⑥ Day Interval	0 — 5	"
⑦ Coord Range	0 — Max	"

* Subtype of Character Data type :- Value
(Alphanumeric + Special Characters)

operation : =

Result - Boolean

(9) Gender :-

Value - M or F

Operation : =

Result type \Rightarrow Boolean.

* String Data Type :-

String are just sequences of character of arbitrary length. A string datatype is made by stringing together a sequence of character.

String — Value — Any sequence of character
operation. Result type
Char in string? Boolean
= Boolean.

* Subtypes of String Data Type :-

(a) Names — Value : Strings with no special characters.

(b) City — Value :

(c) Words — Values: Strings with alphanumeric and .,:;-

All operations allowed (=)

string

(d) Category :-

Values — Can take only one of the following values :

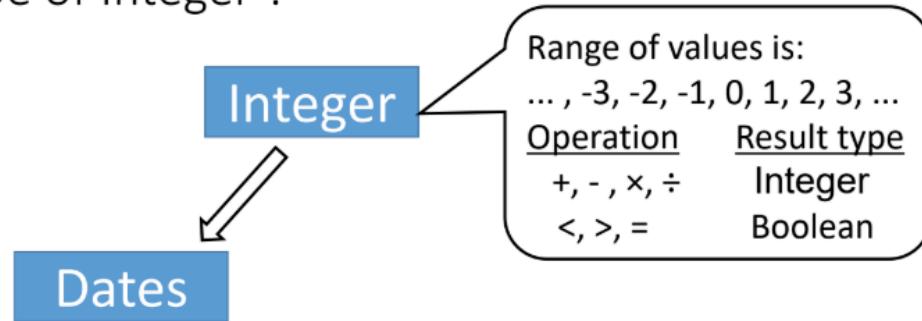
"Noun", "Verb", "Preposition", "Adjective", -

Lecture 1.9 Transformation of Sub-datatypes

Subtypes: transforming values

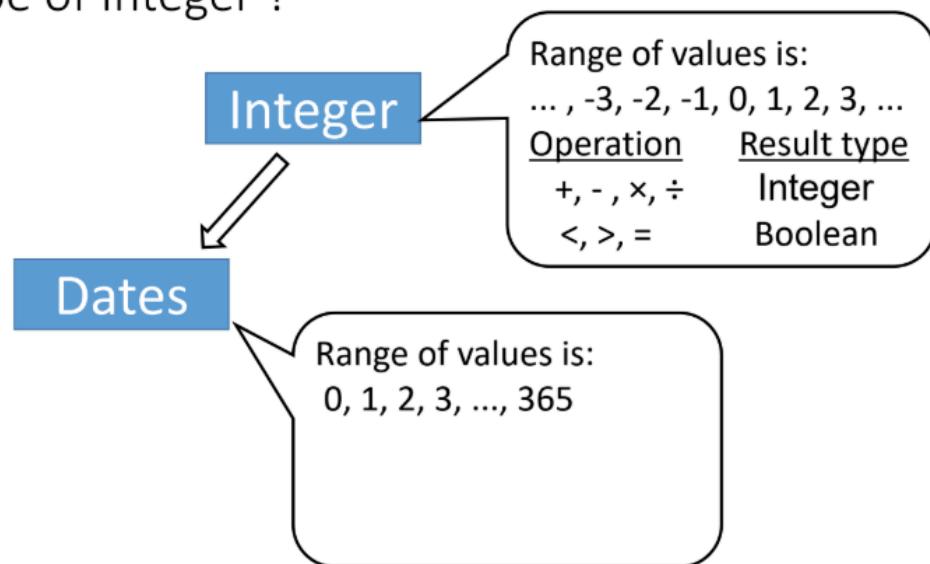
Date : Subtype of Integer

Date: Subtype of Integer ?



2.

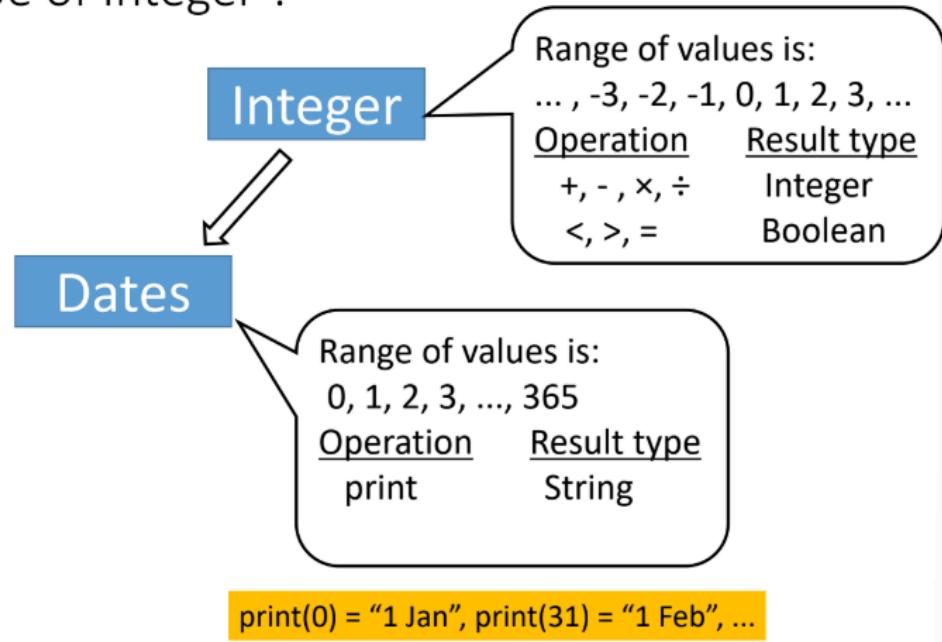
Date: Subtype of Integer ?



Date value is 0 for 1 Jan, 1 for 2 Jan, ..., 30 for 31 Jan, 31 for 1 Feb, ...

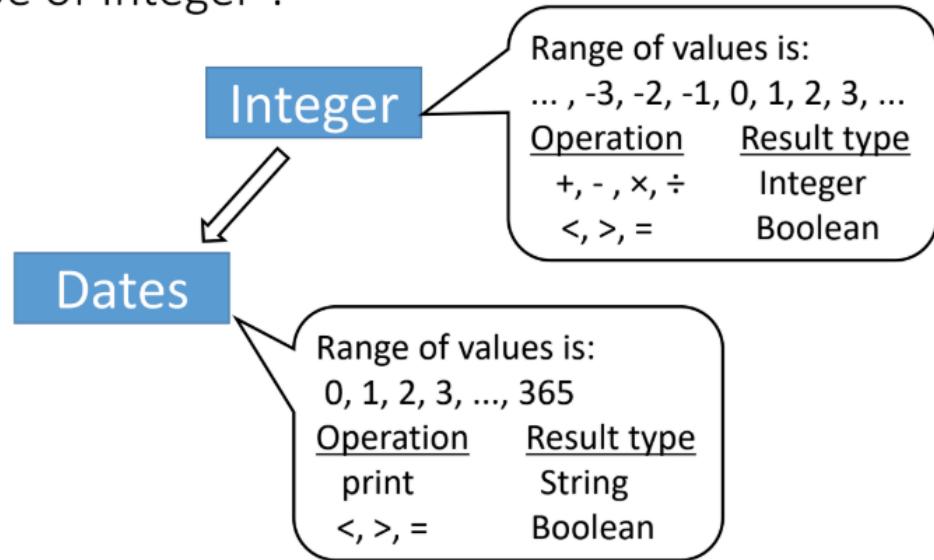
3.

Date: Subtype of Integer ?



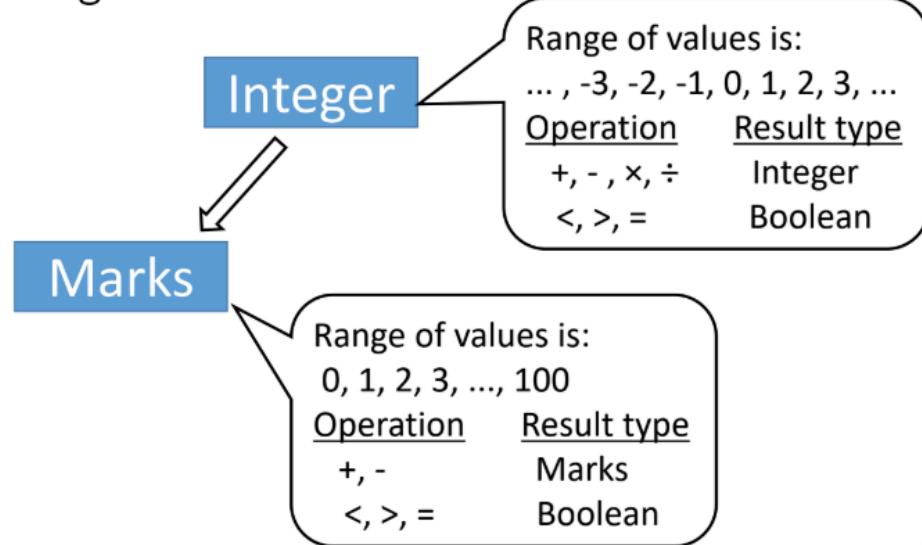
4.

Date: Subtype of Integer ?



5.

Subtypes of Integer: fractional marks

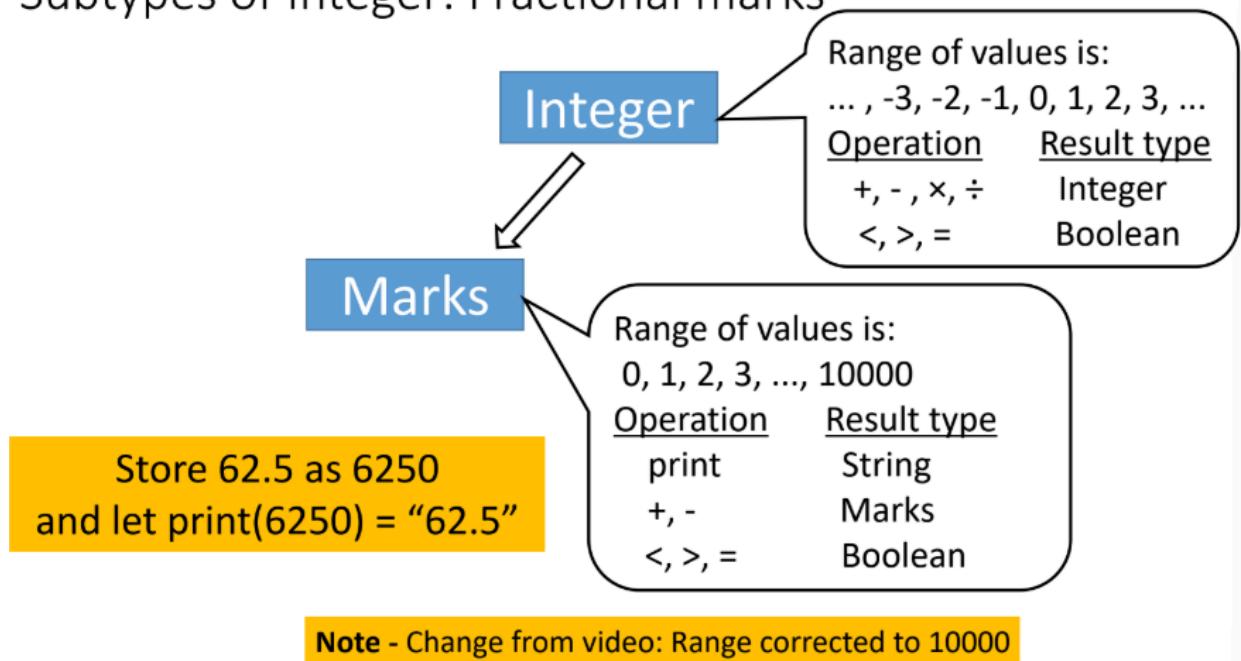


What about fractional marks, e.g. 62.5 ?

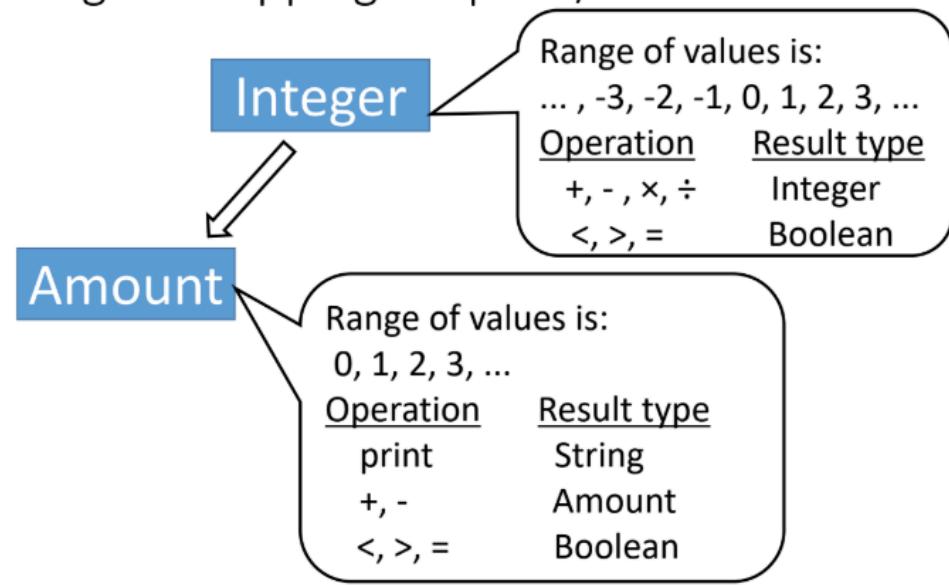
Dealing with fractional values

- Can use another basic type for real numbers - called Float
 - But our values are going to typically be only up to 2 decimal places (e.g. 75.25). So, we have to write constraints for the float values.
- What if we just multiply the fractional number by 100?
 - Then the fractional value with at most 2 decimal places will become an integer !
 - We can do corresponding operations on the integer values (remembering that they have been scaled by 100)
 - And when we finally print the output, we scale the number down and print it

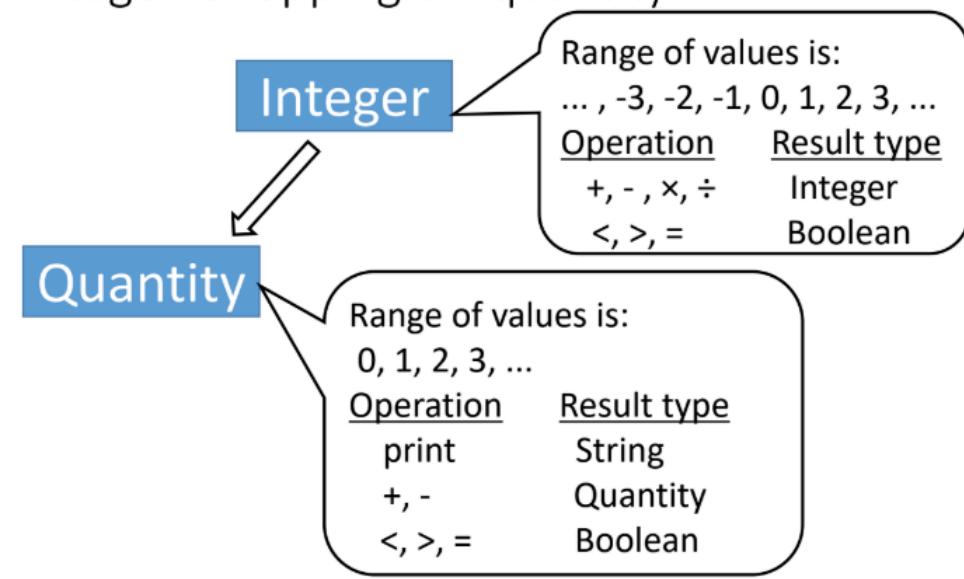
Subtypes of Integer: Fractional marks



Subtypes of Integer: shopping bill price/total



Subtypes of Integer: shopping bill quantity



Store 1.25 as 125 and let print(125) = "1.25"

L. 1.9 - Transformation of Subdatatypes

19/06/2024

Date → String ✓ but in string we can only compare.

- * Date : Subtype of Integer :- Integer

- Date :- Range of Values is :

[0, 1, 2, 3, ..., 365]

Date value is 0 for 1 Jan, 1 for 2 Jan, ..., 31 for 1 Feb.

* Operation :-

• Print

* Result type

String

Print(0) = "1 Jan"
Print(31) = "1 Feb", -

• <, >, =

Boolean

- * Dealing with fractional Values :-

There is yet another datatype that could be used for storing fractional values, which is called float.
But our values are going to typically be only up to 2 decimal places (e.g. 75.25). So, we have to write constraints for the float values.

- What if we just multiply the fractional number by 100?
- Then the fractional value with at most 2 decimal places will become an Integer.
- We can do corresponding operations on the integer values (remembering that they have been scaled by 100).
- And when we finally print the output, we scale the number down and Print it.

+ float $M \times 100$. Integer will we perform all operation which we can perform on Integer.

eg. 62.50 up 62.50 and let $\text{Print}(62.50) = "62.5"$.

Fractional Marks:

Marks :	Range of Value is
	$0, 1, 2, \dots, 1000$
Operation	Result type
Print	String
$+, -$	Marks
$<, >, =$	Boolean

Shopping Bill Price/Total :

Amount :	Range of Values is :
	$0, 1, 2, 3, \dots$
Operation	Result type
Print	String
$+, -$	Amount
$<, >, =$	Boolean

Shopping Bill Quantity :

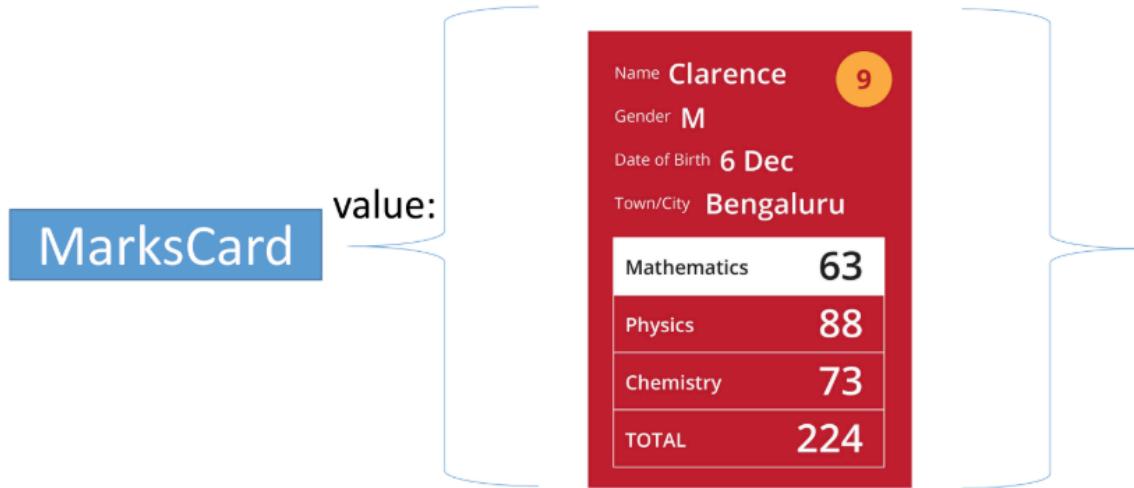
Quantity :	Range of Values is :
	$0, 1, 2, 3, \dots$
Operation	Result type
Print	String
$+, -$	Quantity
$<, >, =$	Boolean

Lecture 1.10 Introduction to complex datatypes

Records and Lists

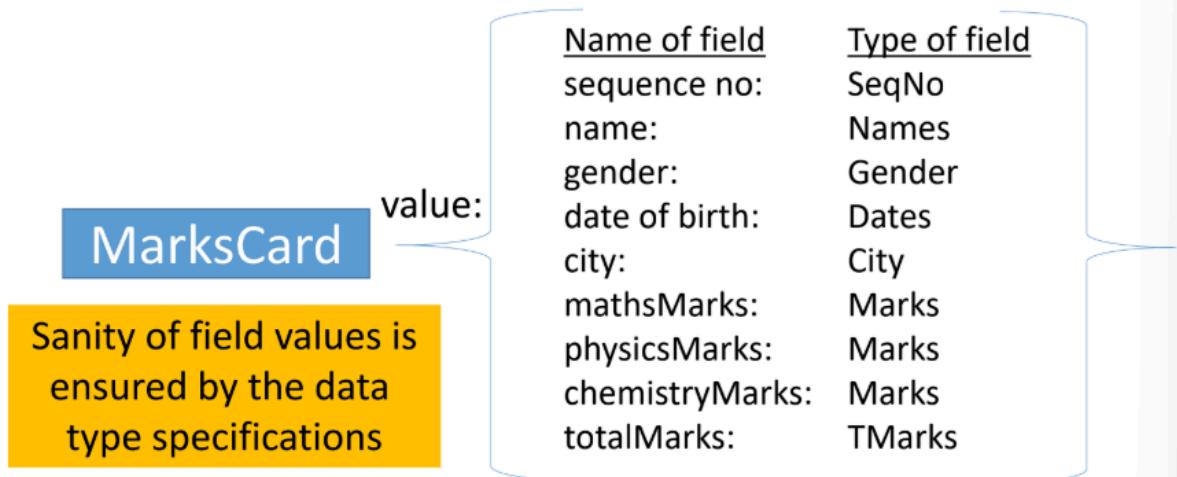
Record (also called struct or tuple)

Data type with multiple fields - each of which has a name and a value



Record (also called struct or tuple)

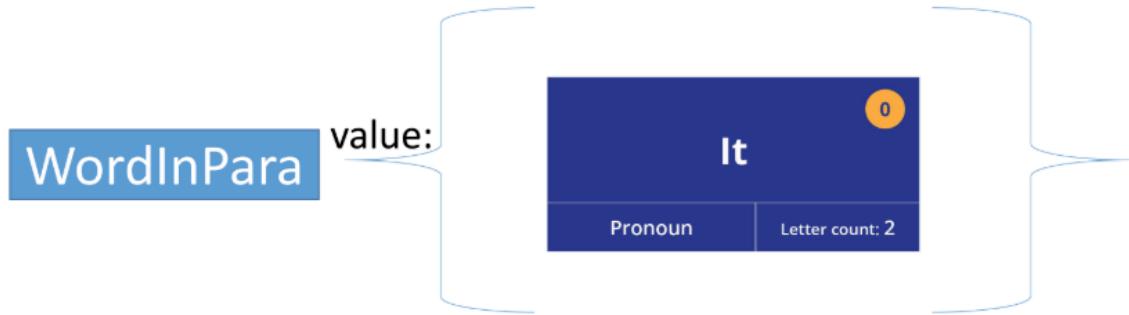
Data type with multiple fields - each of which has a name and a value



Note - Change from video: TMarks for totalMarks is the same as Marks, except range is upto 300

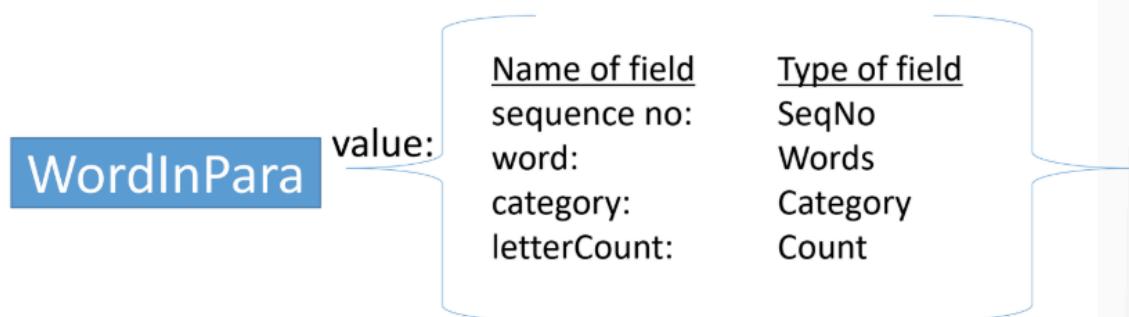
Record (also called struct or tuple)

Data type with multiple fields - each of which has a name and a value



Record (also called struct or tuple)

Data type with multiple fields - each of which has a name and a value



List

- A sequence of data elements (for example a sequence of records)
- MarksCardList - is the data type for our data set of all marks cards
 - Each element in the sequence is of MarksCard Record data type
- ParagraphWordList - is the data type for our word data set
 - Each element in the sequence is of WordInPara Record data type
- ShoppingBillList - data type for the shopping bill data set
 - We need to define the Record data type for a shopping bill

Record (also called struct or tuple)

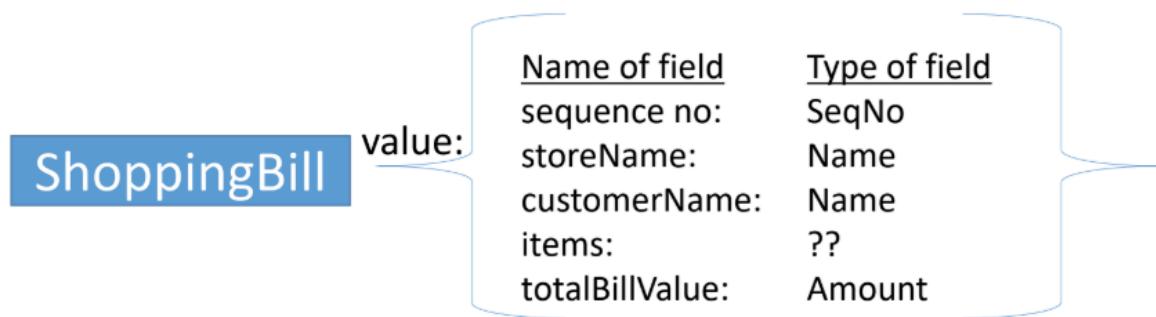
Data type with multiple fields - each of which has a name and a value

ShoppingBill

value:

SV Stores		Srivatsan 1		
Item	Category	Qty	Price	Cost
Carrots	Vegetables/Food	1.5	50	75
Soap	Toiletries	4	32	128
Tomatoes	Vegetables/Food	2	40	80
Bananas	Vegetables/Food	8	8	64
Socks	Footwear/Apparel	3	56	168
Curd	Dairy/Food	0.5	32	16
Milk	Dairy/Food	1.5	24	36
				567

Record data type for shopping bill



Record data type for shopping bill

There is a list of items
in the shopping bill !

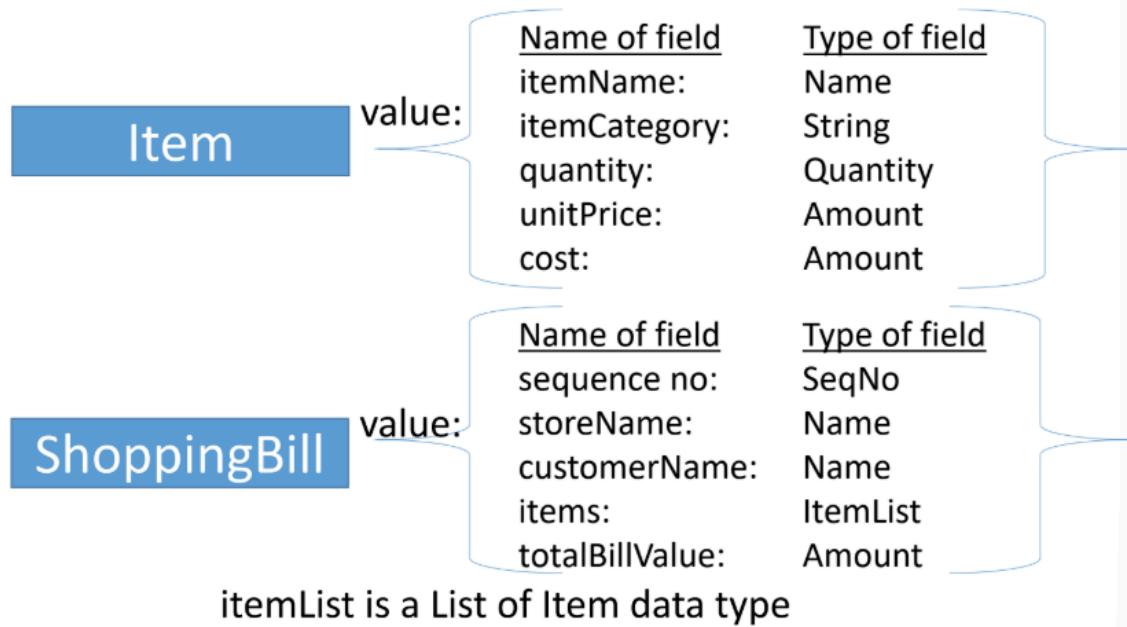
Carrots	Vegetables/Food	1.5	50	75
Soap	Toiletries	4	32	128
Tomatoes	Vegetables/Food	2	40	80
Bananas	Vegetables/Food	8	8	64
Socks	Footwear/Apparel	3	56	168
Curd	Dairy/Food	0.5	32	16
Milk	Dairy/Food	1.5	24	36

ShoppingBill

value:

Name of field	Type of field
sequence no:	SeqNo
storeName:	Name
customerName:	Name
items:	??
totalBillValue:	Amount

Record data type for shopping bill



Summary

- A data type defines the values that the variable can take, and the set of operations that are permitted on it
- Basic data types - Integer, Boolean and Character are needed for our data sets
- Subtypes of a type can be used to further restrict the values and/or operations allowed
- Record type is a collections of named fields, each with the same or different data types
- List type is a sequence of items, all of which are usually of the same data type

CT live session Note : Week 1

Week – 1 CT Course Notes

1. Introduction to Datasets

- A dataset is usually made of records, and each record consists of fields (with names and values).
- A dataset is a structured collection of data.
- **Example datasets:**
 - Scores – records of students' marks
 - Paragraph Words – records of words in a paragraph
 - Shopping Bill – records of items purchased
 - Train - records of train

2. Concept of Variables, Iterators, and Filtering

Variables

- A **variable** is a name given to a value. And this keeps on changing as the computation goes on.

Example:

marks = 95

marks = 100

Extra: Constant vs Changing Variable: Exam Marks Processing

1. Constant Variable

A **constant variable** is one whose value does not change during computation.

Example: MAX_MARKS = 100

- This represents the maximum marks a student can score in the exam.
 - No matter how many students we process, MAX_MARKS will always stay 100.
-

2. Changing Variable

A **changing variable** is one whose value keeps changing as computation proceeds.

Example: highestScore = 0

- Initially set to 0.
 - As we go through the list of student marks, we update highestScore if we find a student with more marks.
-

Iterators

- An **iterator** is a variable that allows traversal(repetition) through elements of a dataset, one by one.

Example: moving through each student's marks in a Scores DataSet.

Filtering

- **Filtering** is applying conditions to select only required records.

Example: Select students with marks > 50.

Description of the Iterator

1. **Initialization step:** Arrange all the cards in an “unseen” pile.
2. **Continue or Exit:** If there are no more cards in the unseen pile, we exit; otherwise, we continue.
3. **Repeat step:** Pick an element from the unseen pile, perform the required operation, and then move it to another “seen” pile.
4. **Go back** to Step 2.

3. Iterations Using Combination of Filtering Conditions

- **Iteration** = repeated traversal over dataset elements.
- **Filtering conditions** may be combined with logical operators:
 - AND, OR, NOT
- **Example:** Select students such that marks > 50 AND gender = 'F'

4. Introduction to Flowcharts

What is a Flowchart?

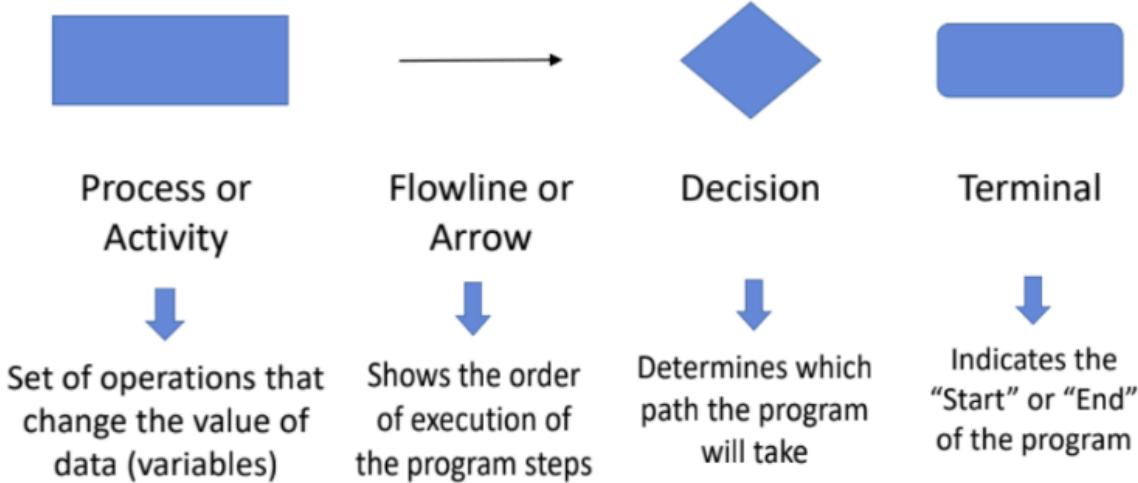
- A flowchart is a graphical representation of an **algorithm**.
- Used by programmers as a program-planning tool to solve problems.
- Consists of symbols connected with arrows to show the flow of data and processing.
- The process of drawing a flowchart is called **flowcharting**.

Commonly Used Symbols :

Symbol	Purpose	Description
Rectangle (Process/Activity)	Represents a set of operations	Used for arithmetic or data processing steps.
Arrow (Flowline)	Shows order of execution	Indicates flow of control.
Diamond (Decision)	Represents a decision point	Determines which path the program will take.
Oval (Terminal)	Start or End	Indicates beginning or end of program.
Parallelogram (Input/Output)	Input/Output	Input from devices or output to screen.
Circle (Connector)	Continuation	Used when flowchart spans multiple pages.

Pictorial View :

Flowcharts



Advantages of Flowcharts

- Easy to communicate the logic of a system.
- Acts as a guide or blueprint during program design.
- Helps in debugging process.
- Makes analysis of programs easier.
- Provides better documentation.

Disadvantages of Flowcharts

- Difficult to draw for large and complex programs.
- No standard for determining level of detail.
- Difficult to reproduce.
- Difficult to modify once drawn.

5. Flowchart Example

Task: Find total maths marks of all students from a record sheet.

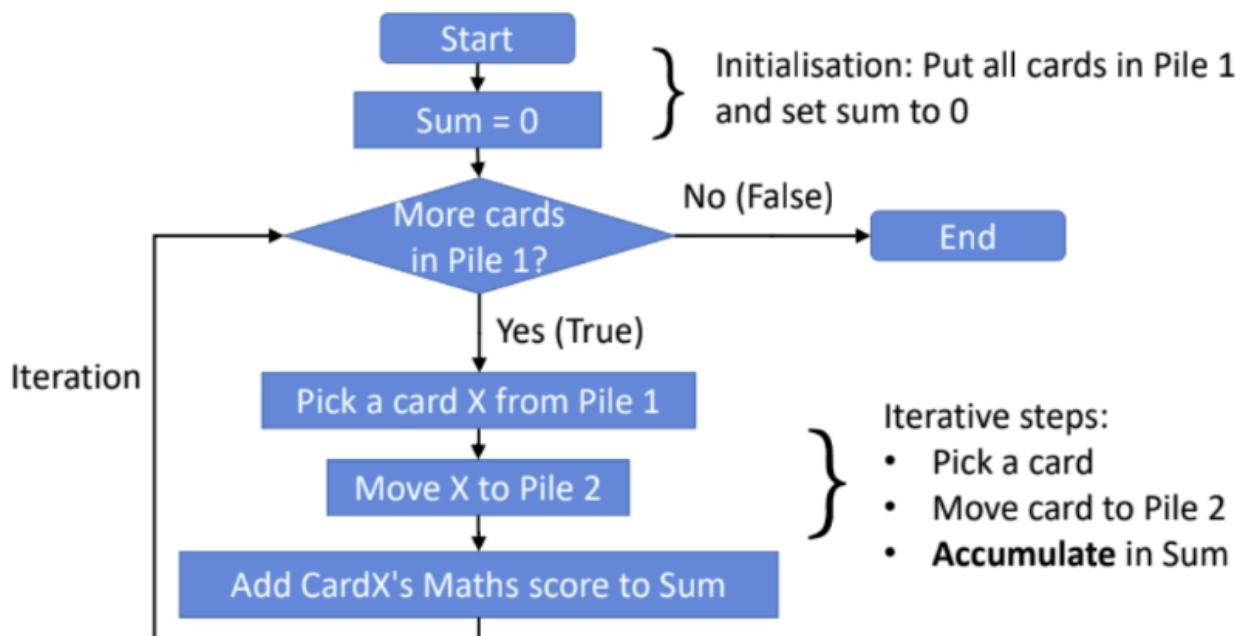
Steps (Flowchart Logic):

1. Start.
2. Initialize Sum = 0.
3. Check if there are cards in Pile 1.
 - o If no → End.
 - o If yes → Pick a card X from Pile 1.
4. Move X to Pile 2.
5. Add card X's Maths score to Sum.
6. Go back to Step 3.

Modifications(You can try):

- If we need only boys' marks.
- If we need only girls' marks.
- If we need boys' and girls' marks separately.

Summary: Flowchart for sum of Maths marks



6. Sanity of Data

- Data is organized into cards, each storing one data item.
- Each card can have multiple elements:
 - Numbers (e.g., marks)
 - Sequence of characters (e.g., name, bill item)

Observations

- Restrictions on values:
 - Marks must lie between 0 and 100.
 - Names cannot contain invalid characters.
 - Constraints on operations:
 - Addition of marks is possible, but multiplication may not make sense.
 - Names can be compared (e.g., one name with another → Boolean True/False).
 - Names cannot be added together.
 -
-

7. Introduction to Datatypes

Concept of Data Types

- By associating a data type with a data element, we define:
- When we specify a variable's type, we describe constraints on the values it can store and the valid operations.
- Datatype defines:
 - The values a variable can take
 - The operations that can be applied

Basic Data Types

1. Boolean
 - Values: True or False
 - Operations: AND, OR, NOT
 - Result type: Boolean
2. Integer
 - Values: ... -3, -2, -1, 0, 1, 2, 3 ...
 - Operations: +, -, *, /, <, >, %
 - Result type: Integer or Boolean

3. Character

- Values: Alphanumeric (A–Z, a–z, 0–9)
- Special characters: , . ; ' " / & % \$ # @ !
- Operations: Limited (comparisons possible)
- Result type: Boolean

4. String

- Values: Any sequence of characters
 - Operations: Membership test (e.g., char in string?)
 - Result type: Boolean
-

8. Subtypes of Basic Datatypes

(a) Marks

- Range: 0, 1, 2, ... 100
 - Operations: =, <, >
 - Result type:
 - Marks
 - Boolean
-

(b) Count

- Range: 0, 1, 2, ...
 - Operations: =, <, >
 - Result type:
 - Count
 - Boolean
 - Arithmetic (* /) does not make sense for count.
-

(c) Gender (Subtype of Character)

- Values: M or F
 - Operations: =
 - Result type: Boolean
-

(d) String Subtypes

- Names → strings without special characters (e.g., city names)
- Words → strings with alphanumeric + . , ;
- Category → limited set of values (e.g., "Noun", "Verb", "Adjective")

9. Transformation of Sub-datatypes

(a) Date (Subtype of Integer)

- Range: 0 ... 365
 - Base values:
 - 0 = 1 Jan
 - 1 = 2 Jan
 - 31 = 1 Feb
 - Operations:
 - `print(date)` → String
 - `<, >` → Boolean
 - Example:
 - `print(0)` = "1 Jan"
 - `print(31)` = "1 Feb"
-

(b) Fractional Marks

- Represented by Float datatype (real numbers).
 - Constraint: usually up to 2 decimal places (e.g., 76.05).
 - Alternative representation:
 - Multiply by 100 when storing → becomes integer
 - Divide by 100 when printing → restores decimal
 - Example:
 - Store 62.5 as 6250
 - `print(6250)` → "62.5"
-

10. Introduction to Complex Datatypes

(a) Record (Struct / Tuple)

- A datatype with multiple fields, each with:
 - Name
 - Value
 - Datatype specification
- Ensures sanity of individual fields.

Example: Item Record

Field	Datatype	Example Value
itemName	String	"Milk"
quantity	Integer	2
unitPrice	Float	25.50
totalPrice	Float	51.00

So, one Item Record looks like :

```
Item = { itemName: "Milk", quantity: 2, unitPrice: 25.50, totalPrice: 51.00 }
```

This is a Record datatype.

(b) List

- A sequence of elements, usually of the same datatype.
- Examples:
 - ScoresList → list of Scores records
 - ParagraphWordsList → list of WordInPara records
 - ShoppingBillList → list of Item records

Here, our shopping bill will be a List of Item Records.

Example: ItemList (List of Items) :

```
ItemList = [
    { itemName: "Milk", quantity: 2, unitPrice: 25.50, totalPrice: 51.00 },
    { itemName: "Bread", quantity: 1, unitPrice: 30.00, totalPrice: 30.00 },
    { itemName: "Eggs", quantity: 12, unitPrice: 6.00, totalPrice: 72.00 }
]
```

Now, let's define the ShoppingBill Record. This is another Record, but one of its fields is itself a List of items.

Field	Datatype	Example Value
billNo	Integer	101
customer	String	"Alice"
date	Date	"12 Sep"
items	List(Item)	ItemList
grandTotal	Float	153.00

Example: ShoppingBill Record

```
ShoppingBill = {  
    billNo: 101,  
    customer: "Alice",  
    date: "12 Sep",  
    items: [  
        { itemName: "Milk", quantity: 2, unitPrice: 25.50, totalPrice:  
51.00 },  
        { itemName: "Bread", quantity: 1, unitPrice: 30.00, totalPrice:  
30.00 },  
        { itemName: "Eggs", quantity: 12, unitPrice: 6.00, totalPrice:  
72.00 }  
    ],  
    grandTotal: 153.00  
}
```

List of Shopping Bills

Finally, a supermarket will have many shopping bills. So we define a ShoppingBillList, which is a **List of ShoppingBill Records**.

Example: ShoppingBillList

```

ShoppingBillList = [
    {
        billNo: 101,
        customer: "Alice",
        date: "12 Sep",
        items: [
            { itemName: "Milk", quantity: 2, unitPrice: 25.50, totalPrice: 51.00 },
            { itemName: "Bread", quantity: 1, unitPrice: 30.00, totalPrice: 30.00 },
            { itemName: "Eggs", quantity: 12, unitPrice: 6.00, totalPrice: 72.00 }
        ],
        grandTotal: 153.00
    },
    {
        billNo: 102,
        customer: "Bob",
        date: "13 Sep",
        items: [
            { itemName: "Rice", quantity: 5, unitPrice: 40.00, totalPrice: 200.00 },
            { itemName: "Oil", quantity: 1, unitPrice: 120.00, totalPrice: 120.00 }
        ],
        grandTotal: 320.00
    }
]

```

Note:

1. Record Datatype = Collection of fields.
 - **Example:** Item with itemName, quantity, unitPrice, totalPrice.
 2. List Datatype = Sequence of elements of the same type.
 - **Example:** ItemList = list of multiple Item Records.
 3. Nested Datatypes:
 - A ShoppingBill Record contains a field items which is a List of Item Records.
 - A ShoppingBillList is a list of ShoppingBill Records.
-

Summary :

- A datatype defines valid values and operations.
- Basic datatypes: Integer, Boolean, Character, String.
- Subtypes restrict ranges and permitted operations (e.g., Marks, Count, Gender).
- Transformations allow mapping between integer/float and human-readable formats (e.g., Date, Fractional marks).
- Record: collection of named fields (possibly different datatypes or sometimes same).
- List: sequence of elements of the same datatype, enabling representation of datasets.

PA : QN 8,9,10,11

For 9th check Qny of dustpan

For 11th check conditions taking different numbers

GA - QN : 2,5,7,8

2 - Jo bola vo karo <250 > 250 means ignore 250