

Lecture 1, 8 August 2023

Madhavan Mukund

<https://www.cmi.ac.in/~madhavan>

Programming and Data Structures with Python

What is programming?

- Writing systematic procedures in precise notation
 - Systematic procedure: **algorithm**
 - Precise notation: **programming language**
- **Example:** Prepare a classroom for a seminar by a guest speaker
 - Things to do: arrange chairs, check projector, check audio/video, turn on a/c early, ...
 - Need to instruct support staff to do this task
- Nature of instructions varies according to who is doing the job
 - Outsource: Just provide the time of the talk and the expected audience size.
 - Experienced staff: High-level checklist, need not describe each step explicitly
 - Inexperienced staff: Each step needs detailed instructions
 - Arranging chairs: arrange m rows of chairs, k chairs per row, leave aisles in between to walk to the back, ...

Placement data

Year	Programme	Domain	Package
2017-18	B.Sc	CS	1800000
2017-18	M.Sc Appl. Math	CS	1170000
2018-19	B.Sc	Banking-Finance	700000
2018-19	M.Sc Appl. Math	Manufacturing	730000
2019-20	M.Sc Data Science	CS	2000000
2019-20	M.Sc Data Science	Analytics	1700000
2019-20	M.Sc Data Science	Banking-Finance	1344000
2019-20	M.Sc Comp. Science	Banking-Finance	1335827
2019-20	M.Sc Comp. Science	Logistics	1000000
2020-21	M.Sc Data Science	Analytics	1680000
2020-21	M.Sc Data Science	CS	1150000
2020-21	Ph.D Physics	Banking-Finance	3000000

Questions

- How many students were placed overall?
- What is the maximum pay package?
- What is the average pay package?
- How many students have above average pay packages?
- Was anyone placed in Logistics?
- How many students were placed in Banking-Finance?
- Which domain has maximum placements?
- Which domain has the best average pay package?

Typical questions

- How many students were placed overall?
 - How is our table made available to us?
 - A **list** of rows, each is **tuple** of columns
 - [row-1,row-2,...,row-N]
 - row-j is (year, programme, domain, pay package)
 - Run through all the rows from beginning to end — **iteration**
 - Maintain a counter, **variable count**
 - Initialize to 0
 - Increment **count** with each row
 - Report value of **count** at the end of the iteration

Year	Programme	Domain	Package
2017-18	B.Sc	CS	1800000
2017-18	M.Sc Appl. Math	Banking-Finance	1490000
2017-18	M.Sc Appl. Math	Logistics	1250000
2017-18	M.Sc Appl. Math	CS	2000000
2018-19	M.Sc Appl. Math	Manufacturing	730000
2019-20	M.Sc Data Science	Analytics	1700000
2019-20	M.Sc Data Science	Banking-Finance	1344000
2019-20	M.Sc Data Science	Banking-Finance	1335827
2019-20	M.Sc Data Science	CS	1301968
2019-20	M.Sc Comp. Science	Logistics	1000000
2020-21	M.Sc Comp. Science	Banking-Finance	1500000
2020-21	M.Sc Data Science	Analytics	900000
2020-21	Ph.D Comp. Science	Banking-Finance	1454545
2020-21	M.Sc Data Science	Analytics	1300000
2020-21	M.Sc Data Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Analytics	1680000
2020-21	M.Sc Data Science	Banking-Finance	1360000
2020-21	M.Sc Data Science	Banking-Finance	1842632
2020-21	M.Sc Data Science	CS	1150000
2020-21	M.Sc Data Science	Analytics	850000
2020-21	M.Sc Comp. Science	Analytics	850000
2020-21	M.Sc Data Science	CS	1200000
2020-21	M.Sc Data Science	Analytics	1608000
2020-21	M.Sc Data Science	CS	1100000
2020-21	M.Sc Data Science	Banking-Finance	1550000
2020-21	M.Sc Comp. Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Banking-Finance	1200000
2020-21	Ph.D Physics	Banking-Finance	3000000

Typical questions

- What is the maximum pay package?
 - Again iterate through the rows
 - Maintain a variable `max` — the maximum salary we have seen so far
 - Initialize to `0` — lower bound, no pay package is negative
 - Can also initialize `max` to pay package in first row — maximum is not meaningful for an empty table
 - For each row, if current pay exceeds `max`, update `max` to the current pay
 - At the end of the iteration, `max` is the largest pay package

Year	Programme	Domain	Package
2017-18	B.Sc	CS	1800000
2017-18	M.Sc Appl. Math	Banking-Finance	1490000
2017-18	M.Sc Appl. Math	Logistics	1250000
2017-18	M.Sc Appl. Math	CS	2000000
2018-19	M.Sc Appl. Math	Manufacturing	730000
2019-20	M.Sc Data Science	Analytics	1700000
2019-20	M.Sc Data Science	Banking-Finance	1344000
2019-20	M.Sc Data Science	Banking-Finance	1335827
2019-20	M.Sc Data Science	CS	1301968
2019-20	M.Sc Comp. Science	Logistics	1000000
2020-21	M.Sc Comp. Science	Banking-Finance	1500000
2020-21	M.Sc Data Science	Analytics	900000
2020-21	Ph.D Comp. Science	Banking-Finance	1454545
2020-21	M.Sc Data Science	Analytics	1300000
2020-21	M.Sc Data Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Analytics	1680000
2020-21	M.Sc Data Science	Banking-Finance	1360000
2020-21	M.Sc Data Science	Banking-Finance	1842632
2020-21	M.Sc Data Science	CS	1150000
2020-21	M.Sc Data Science	Analytics	850000
2020-21	M.Sc Comp. Science	Analytics	850000
2020-21	M.Sc Data Science	CS	1200000
2020-21	M.Sc Data Science	Analytics	1608000
2020-21	M.Sc Data Science	CS	1100000
2020-21	M.Sc Data Science	Banking-Finance	1550000
2020-21	M.Sc Comp. Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Banking-Finance	1200000
2020-21	Ph.D Physics	Banking-Finance	3000000

Typical questions

- What is the average pay package?
 - Need overall count and total sum of pay packages
 - Already know how to iterate and count
 - Total sum: iterate over rows and update variable `sum`
 - Initialize to 0
 - For each row, add current pay to `sum`
 - Average is `sum/count`
 - Naively, two iterations, one for `count` and another for `sum`
 - Collapse into a single iteration, update `count` and `sum` with each row

Year	Programme	Domain	Package
2017-18	B.Sc	CS	1800000
2017-18	M.Sc Appl. Math	Banking-Finance	1490000
2017-18	M.Sc Appl. Math	Logistics	1250000
2017-18	M.Sc Appl. Math	CS	2000000
2018-19	M.Sc Appl. Math	Manufacturing	730000
2019-20	M.Sc Data Science	Analytics	1700000
2019-20	M.Sc Data Science	Banking-Finance	1344000
2019-20	M.Sc Data Science	Banking-Finance	1335827
2019-20	M.Sc Data Science	CS	1301968
2019-20	M.Sc Comp. Science	Logistics	1000000
2020-21	M.Sc Comp. Science	Banking-Finance	1500000
2020-21	M.Sc Data Science	Analytics	900000
2020-21	Ph.D Comp. Science	Banking-Finance	1454545
2020-21	M.Sc Data Science	Analytics	1300000
2020-21	M.Sc Data Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Analytics	1680000
2020-21	M.Sc Data Science	Banking-Finance	1360000
2020-21	M.Sc Data Science	Banking-Finance	1842632
2020-21	M.Sc Data Science	CS	1150000
2020-21	M.Sc Data Science	Analytics	850000
2020-21	M.Sc Comp. Science	Analytics	850000
2020-21	M.Sc Data Science	CS	1200000
2020-21	M.Sc Data Science	Analytics	1608000
2020-21	M.Sc Data Science	CS	1100000
2020-21	M.Sc Data Science	Banking-Finance	1550000
2020-21	M.Sc Comp. Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Banking-Finance	1200000
2020-21	Ph.D Physics	Banking-Finance	3000000

Typical questions

- How many students have above average pay packages?
 - First iteration to compute average
 - Second iteration to count students above average
 - Maintain variable `aboveaverage`
 - Initialize to 0
 - For each row, increment `aboveaverage` if current pay is above the average
 - Filtered update

Year	Programme	Domain	Package
2017-18	B.Sc	CS	1800000
2017-18	M.Sc Appl. Math	Banking-Finance	1490000
2017-18	M.Sc Appl. Math	Logistics	1250000
2017-18	M.Sc Appl. Math	CS	2000000
2018-19	M.Sc Appl. Math	Manufacturing	730000
2019-20	M.Sc Data Science	Analytics	1700000
2019-20	M.Sc Data Science	Banking-Finance	1344000
2019-20	M.Sc Data Science	Banking-Finance	1335827
2019-20	M.Sc Data Science	CS	1301968
2019-20	M.Sc Comp. Science	Logistics	1000000
2020-21	M.Sc Comp. Science	Banking-Finance	1500000
2020-21	M.Sc Data Science	Analytics	900000
2020-21	Ph.D Comp. Science	Banking-Finance	1454545
2020-21	M.Sc Data Science	Analytics	1300000
2020-21	M.Sc Data Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Analytics	1680000
2020-21	M.Sc Data Science	Banking-Finance	1360000
2020-21	M.Sc Data Science	Banking-Finance	1842632
2020-21	M.Sc Data Science	CS	1150000
2020-21	M.Sc Data Science	Analytics	850000
2020-21	M.Sc Comp. Science	Analytics	850000
2020-21	M.Sc Data Science	CS	1200000
2020-21	M.Sc Data Science	Analytics	1608000
2020-21	M.Sc Data Science	CS	1100000
2020-21	M.Sc Data Science	Banking-Finance	1550000
2020-21	M.Sc Comp. Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Banking-Finance	1200000
2020-21	Ph.D Physics	Banking-Finance	3000000

Typical questions

- Was anyone placed in Logistics?
 - Iterate through the rows
 - Check if domain is Logistics
 - Two ways to exit the iteration
 - Found a row with domain Logistics — no need to examine any more rows
 - All rows are exhausted without finding domain Logistics
- Conditional iteration

Year	Programme	Domain	Package
2017-18	B.Sc	CS	1800000
2017-18	M.Sc Appl. Math	Banking-Finance	1490000
2017-18	M.Sc Appl. Math	Logistics	1250000
2017-18	M.Sc Appl. Math	CS	2000000
2018-19	M.Sc Appl. Math	Manufacturing	730000
2019-20	M.Sc Data Science	Analytics	1700000
2019-20	M.Sc Data Science	Banking-Finance	1344000
2019-20	M.Sc Data Science	Banking-Finance	1335827
2019-20	M.Sc Data Science	CS	1301968
2019-20	M.Sc Comp. Science	Logistics	1000000
2020-21	M.Sc Comp. Science	Banking-Finance	1500000
2020-21	M.Sc Data Science	Analytics	900000
2020-21	Ph.D Comp. Science	Banking-Finance	1454545
2020-21	M.Sc Data Science	Analytics	1300000
2020-21	M.Sc Data Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Analytics	1680000
2020-21	M.Sc Data Science	Banking-Finance	1360000
2020-21	M.Sc Data Science	Banking-Finance	1842632
2020-21	M.Sc Data Science	CS	1150000
2020-21	M.Sc Data Science	Analytics	850000
2020-21	M.Sc Comp. Science	Analytics	850000
2020-21	M.Sc Data Science	CS	1200000
2020-21	M.Sc Data Science	Analytics	1608000
2020-21	M.Sc Data Science	CS	1100000
2020-21	M.Sc Data Science	Banking-Finance	1550000
2020-21	M.Sc Comp. Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Banking-Finance	1200000
2020-21	Ph.D Physics	Banking-Finance	3000000

Typical questions

- How many students were placed in Banking-Finance?
 - Iterate through the rows
 - Maintain a counter
 - Increment only if domain is Banking-Finance
 - Again **filtered** iteration

Year	Programme	Domain	Package
2017-18	B.Sc	CS	1800000
2017-18	M.Sc Appl. Math	Banking-Finance	1490000
2017-18	M.Sc Appl. Math	Logistics	1250000
2017-18	M.Sc Appl. Math	CS	2000000
2018-19	M.Sc Appl. Math	Manufacturing	730000
2019-20	M.Sc Data Science	Analytics	1700000
2019-20	M.Sc Data Science	Banking-Finance	1344000
2019-20	M.Sc Data Science	Banking-Finance	1335827
2019-20	M.Sc Data Science	CS	1301968
2019-20	M.Sc Comp. Science	Logistics	1000000
2020-21	M.Sc Comp. Science	Banking-Finance	1500000
2020-21	M.Sc Data Science	Analytics	900000
2020-21	Ph.D Comp. Science	Banking-Finance	1454545
2020-21	M.Sc Data Science	Analytics	1300000
2020-21	M.Sc Data Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Analytics	1680000
2020-21	M.Sc Data Science	Banking-Finance	1360000
2020-21	M.Sc Data Science	Banking-Finance	1842632
2020-21	M.Sc Data Science	CS	1150000
2020-21	M.Sc Data Science	Analytics	850000
2020-21	M.Sc Comp. Science	Analytics	850000
2020-21	M.Sc Data Science	CS	1200000
2020-21	M.Sc Data Science	Analytics	1608000
2020-21	M.Sc Data Science	CS	1100000
2020-21	M.Sc Data Science	Banking-Finance	1550000
2020-21	M.Sc Comp. Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Banking-Finance	1200000
2020-21	Ph.D Physics	Banking-Finance	3000000

Typical questions

- Which domain has maximum placements?
 - Count placements for each domain and take the max
 - One counter per domain — but we don't know the domains in advance!
 - Maintain set of counters as a function, mapping domains to counts
 - Initialize `count` to empty function
 - For each row
 - If the domain `d` has not been seen before, set `count(d) = 1`
 - otherwise, increment `count(d)`
 - `count` is a collection of (key,value) pairs — called a **dictionary**
 - Iterate over keys of `count` to compute **max**

Year	Programme	Domain	Package
2017-18	B.Sc	CS	1800000
2017-18	M.Sc Appl. Math	Banking-Finance	1490000
2017-18	M.Sc Appl. Math	Logistics	1250000
2017-18	M.Sc Appl. Math	CS	2000000
2018-19	M.Sc Appl. Math	Manufacturing	730000
2019-20	M.Sc Data Science	Analytics	1700000
2019-20	M.Sc Data Science	Banking-Finance	1344000
2019-20	M.Sc Data Science	Banking-Finance	1335827
2019-20	M.Sc Data Science	CS	1301968
2019-20	M.Sc Comp. Science	Logistics	1000000
2020-21	M.Sc Comp. Science	Banking-Finance	1500000
2020-21	M.Sc Data Science	Analytics	900000
2020-21	Ph.D Comp. Science	Banking-Finance	1454545
2020-21	M.Sc Data Science	Analytics	1300000
2020-21	M.Sc Data Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Analytics	1680000
2020-21	M.Sc Data Science	Banking-Finance	1360000
2020-21	M.Sc Data Science	Banking-Finance	1842632
2020-21	M.Sc Data Science	CS	1150000
2020-21	M.Sc Data Science	Analytics	850000
2020-21	M.Sc Comp. Science	Analytics	850000
2020-21	M.Sc Data Science	CS	1200000
2020-21	M.Sc Data Science	Analytics	1608000
2020-21	M.Sc Data Science	CS	1100000
2020-21	M.Sc Data Science	Banking-Finance	1550000
2020-21	M.Sc Comp. Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Banking-Finance	1200000
2020-21	Ph.D Physics	Banking-Finance	3000000

$f(1) = 1$
 $f(2) = 4$
⋮

$f(10) = 100$

$\{(1, 1), (2, 4), (3, 9), \dots, (10, 100)\}$

$(CS, -)$, $(Man, -)$, $(B-F, -)$ —
Key Value

Typical questions

- Which domain has the best average pay package?
 - Use one dictionary to maintain number of students for each domain
 - Another dictionary to maintain total pay for each domain
 - Iterate over the keys of these two dictionaries to construct a dictionary with average pay for each domain
 - Iterate over the keys of dictionary with average pay to compute domain with max average pay
 - Keep track of and update both name of max domain and max pay

Year	Programme	Domain	Package
2017-18	B.Sc	CS	1800000
2017-18	M.Sc Appl. Math	Banking-Finance	1490000
2017-18	M.Sc Appl. Math	Logistics	1250000
2017-18	M.Sc Appl. Math	CS	2000000
2018-19	M.Sc Appl. Math	Manufacturing	730000
2019-20	M.Sc Data Science	Analytics	1700000
2019-20	M.Sc Data Science	Banking-Finance	1344000
2019-20	M.Sc Data Science	Banking-Finance	1335827
2019-20	M.Sc Data Science	CS	1301968
2019-20	M.Sc Comp. Science	Logistics	1000000
2020-21	M.Sc Comp. Science	Banking-Finance	1500000
2020-21	M.Sc Data Science	Analytics	900000
2020-21	Ph.D Comp. Science	Banking-Finance	1454545
2020-21	M.Sc Data Science	Analytics	1300000
2020-21	M.Sc Data Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Analytics	1680000
2020-21	M.Sc Data Science	Banking-Finance	1360000
2020-21	M.Sc Data Science	Banking-Finance	1842632
2020-21	M.Sc Data Science	CS	1150000
2020-21	M.Sc Data Science	Analytics	850000
2020-21	M.Sc Comp. Science	Analytics	850000
2020-21	M.Sc Data Science	CS	1200000
2020-21	M.Sc Data Science	Analytics	1608000
2020-21	M.Sc Data Science	CS	1100000
2020-21	M.Sc Data Science	Banking-Finance	1550000
2020-21	M.Sc Comp. Science	Banking-Finance	1350000
2020-21	M.Sc Data Science	Banking-Finance	1200000
2020-21	Ph.D Physics	Banking-Finance	3000000

Dictionary 1

Domain Count

CS → 5

Man. → 7

B.P → 22

Dictionary 2

Domain Total

CS → 75,00,000

CS \rightarrow (CScount, CSval)
 +1 +Package

Summary

- Programming involves computing with information different types
- Variables hold intermediate values — data types
- Collections of values — lists, tuples, dictionaries
- Processing collections — iteration, conditional termination, filtering