Lecture 1, 7 August 2025

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, ...

Programming for data science — IPL 2024

City	Team 1	Team 2	Toss win	Match win	Run target
Chennai	RCB	CSK	RCB	CSK	174
Mohali	DC	PK	PK	PK	175
Kolkata	KKR	SRH	SRH	KKR	209
Jaipur	RR	LSG	RR	RR	194
Ahmedabad	GT	MI	MI	GT	169
Bengaluru	PK	RCB	RCB	RCB	177
Chennai	CSK	GT	GT	CSK	207
Hyderabad	SRH	MI	MI	SRH	278
Jaipur	RR	DC	DC	RR	186
Bengaluru	RCB	KKR	KKR	KKR	183
Lucknow	LSG	PK	LSG	LSG	200
Ahmedabad	SRH	GT	SRH	GT	163
Visakhapatnam	DC	CSK	DC	DC	192
Mumbai	MI	RR	RR	RR	126
Bengaluru	LSG	RCB	RCB	LSG	182
...	...	...	...	...	...
Chennai	SRH	RR	RR	SRH	176
Chennai	SRH	KKR	SRH	KKR	114

Programming for data science

Questions

- How many matches were played?
- What was the maximum run target?
- What was the average run target?
- How many matches had above average run targets?
- How many cities were venues?
- Which team played as Team 1 at maximum number of venues?
- Is winning the toss an advantage?

Typical questions

- How many matches were played?
  - 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 `(City, Team 1, Team 2, Toss winner, Match winner, Run target)`
  - 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

City	Team 1	Team 2	Toss winner	Match winner	Run target
Chennai	RCB	CSK	RCB	CSK	174
Mohali	DC	PK	PK	PK	175
Kolkata	KKR	SRH	SRH	KKR	209
Jaipur	RR	LSG	RR	RR	194
Ahmedabad	GT	MI	MI	GT	169
Bengaluru	PK	RCB	RCB	RCB	177
Chennai	CSK	GT	GT	CSK	207
Hyderabad	SRH	MI	MI	SRH	278
Jaipur	RR	DC	DC	RR	186
Bengaluru	RCB	KKR	KKR	KKR	183
Lucknow	LSG	PK	LSG	LSG	200
Ahmedabad	SRH	GT	SRH	GT	163
Visakhapatnam	DC	CSK	DC	DC	192
Mumbai	MI	RR	RR	RR	126
Bengaluru	LSG	RCB	RCB	LSG	182
Visakhapatnam	KKR	DC	KKR	KKR	273
Ahmedabad	GT	PK	PK	PK	200
Hyderabad	CSK	SRH	SRH	SRH	166
Jaipur	RCB	RR	RR	RR	184
Mumbai	MI	DC	DC	MI	235
Lucknow	LSG	GT	LSG	LSG	164
Chennai	KKR	CSK	CSK	CSK	138
Mohali	SRH	PK	PK	SRH	183
Jaipur	RR	GT	GT	GT	197
Mumbai	RCB	MI	MI	MI	197
...	...	...	...	...	...
Chennai	SRH	RR	RR	SRH	176
Chennai	SRH	KKR	SRH	KKR	114

Typical questions

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

City	Team 1	Team 2	Toss winner	Match winner	Run target
Chennai	RCB	CSK	RCB	CSK	174
Mohali	DC	PK	PK	PK	175
Kolkata	KKR	SRH	SRH	KKR	209
Jaipur	RR	LSG	RR	RR	194
Ahmedabad	GT	MI	MI	GT	169
Bengaluru	PK	RCB	RCB	RCB	177
Chennai	CSK	GT	GT	CSK	207
Hyderabad	SRH	MI	MI	SRH	278
Jaipur	RR	DC	DC	RR	186
Bengaluru	RCB	KKR	KKR	KKR	183
Lucknow	LSG	PK	LSG	LSG	200
Ahmedabad	SRH	GT	SRH	GT	163
Visakhapatnam	DC	CSK	DC	DC	192
Mumbai	MI	RR	RR	RR	126
Bengaluru	LSG	RCB	RCB	LSG	182
Visakhapatnam	KKR	DC	KKR	KKR	273
Ahmedabad	GT	PK	PK	PK	200
Hyderabad	CSK	SRH	SRH	SRH	166
Jaipur	RCB	RR	RR	RR	184
Mumbai	MI	DC	DC	MI	235
Lucknow	LSG	GT	LSG	LSG	164
Chennai	KKR	CSK	CSK	CSK	138
Mohali	SRH	PK	PK	SRH	183
Jaipur	RR	GT	GT	GT	197
Mumbai	RCB	MI	MI	MI	197
...	...	...	...	...	...
Chennai	SRH	RR	RR	SRH	176
Chennai	SRH	KKR	SRH	KKR	114

Typical questions

- What was the average run target?
  - Need overall count and total sum of run targets
  - Already know how to iterate and count
  - Total sum: iterate over rows and update variable `sum`
    - Initialize to `0`
    - For each row, add current target 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

City	Team 1	Team 2	Toss winner	Match winner	Run target
Chennai	RCB	CSK	RCB	CSK	174
Mohali	DC	PK	PK	PK	175
Kolkata	KKR	SRH	SRH	KKR	209
Jaipur	RR	LSG	RR	RR	194
Ahmedabad	GT	MI	MI	GT	169
Bengaluru	PK	RCB	RCB	RCB	177
Chennai	CSK	GT	GT	CSK	207
Hyderabad	SRH	MI	MI	SRH	278
Jaipur	RR	DC	DC	RR	186
Bengaluru	RCB	KKR	KKR	KKR	183
Lucknow	LSG	PK	LSG	LSG	200
Ahmedabad	SRH	GT	SRH	GT	163
Visakhapatnam	DC	CSK	DC	DC	192
Mumbai	MI	RR	RR	RR	126
Bengaluru	LSG	RCB	RCB	LSG	182
Visakhapatnam	KKR	DC	KKR	KKR	273
Ahmedabad	GT	PK	PK	PK	200
Hyderabad	CSK	SRH	SRH	SRH	166
Jaipur	RCB	RR	RR	RR	184
Mumbai	MI	DC	DC	MI	235
Lucknow	LSG	GT	LSG	LSG	164
Chennai	KKR	CSK	CSK	CSK	138
Mohali	SRH	PK	PK	SRH	183
Jaipur	RR	GT	GT	GT	197
Mumbai	RCB	MI	MI	MI	197
...	...	...	...	...	...
Chennai	SRH	RR	RR	SRH	176
Chennai	SRH	KKR	SRH	KKR	114

Typical questions

- How many matches had above average run targets?
  - First iteration to compute average
  - Second iteration to count matches above average
  - Maintain variable `aboveaverage`
    - Initialize to 0
    - For each row, increment `aboveaverage` if current target is above the average
  - Filtered update

City	Team 1	Team 2	Toss winner	Match winner	Run target
Chennai	RCB	CSK	RCB	CSK	174
Mohali	DC	PK	PK	PK	175
Kolkata	KKR	SRH	SRH	KKR	209
Jaipur	RR	LSG	RR	RR	194
Ahmedabad	GT	MI	MI	GT	169
Bengaluru	PK	RCB	RCB	RCB	177
Chennai	CSK	GT	GT	CSK	207
Hyderabad	SRH	MI	MI	SRH	278
Jaipur	RR	DC	DC	RR	186
Bengaluru	RCB	KKR	KKR	KKR	183
Lucknow	LSG	PK	LSG	LSG	200
Ahmedabad	SRH	GT	SRH	GT	163
Visakhapatnam	DC	CSK	DC	DC	192
Mumbai	MI	RR	RR	RR	126
Bengaluru	LSG	RCB	RCB	LSG	182
Visakhapatnam	KKR	DC	KKR	KKR	273
Ahmedabad	GT	PK	PK	PK	200
Hyderabad	CSK	SRH	SRH	SRH	166
Jaipur	RCB	RR	RR	RR	184
Mumbai	MI	DC	DC	MI	235
Lucknow	LSG	GT	LSG	LSG	164
Chennai	KKR	CSK	CSK	CSK	138
Mohali	SRH	PK	PK	SRH	183
Jaipur	RR	GT	GT	GT	197
Mumbai	RCB	MI	MI	MI	197
...	...	...	...	...	...
Chennai	SRH	RR	RR	SRH	176
Chennai	SRH	KKR	SRH	KKR	114

Typical questions

- How many cities were venues?
  - Maintain a list of cities
  - Check if current city is already in the list. If not, add it.
  - Count items in the list of cities
  - Can we do better?

City	Team 1	Team 2	Toss winner	Match winner	Run target
Chennai	RCB	CSK	RCB	CSK	174
Mohali	DC	PK	PK	PK	175
Kolkata	KKR	SRH	SRH	KKR	209
Jaipur	RR	LSG	RR	RR	194
Ahmedabad	GT	MI	MI	GT	169
Bengaluru	PK	RCB	RCB	RCB	177
Chennai	CSK	GT	GT	CSK	207
Hyderabad	SRH	MI	MI	SRH	278
Jaipur	RR	DC	DC	RR	186
Bengaluru	RCB	KKR	KKR	KKR	183
Lucknow	LSG	PK	LSG	LSG	200
Ahmedabad	SRH	GT	SRH	GT	163
Visakhapatnam	DC	CSK	DC	DC	192
Mumbai	MI	RR	RR	RR	126
Bengaluru	LSG	RCB	RCB	LSG	182
Visakhapatnam	KKR	DC	KKR	KKR	273
Ahmedabad	GT	PK	PK	PK	200
Hyderabad	CSK	SRH	SRH	SRH	166
Jaipur	RCB	RR	RR	RR	184
Mumbai	MI	DC	DC	MI	235
Lucknow	LSG	GT	LSG	LSG	164
Chennai	KKR	CSK	CSK	CSK	138
Mohali	SRH	PK	PK	SRH	183
Jaipur	RR	GT	GT	GT	197
Mumbai	RCB	MI	MI	MI	197
...	...	...	...	...	...
Chennai	SRH	RR	RR	SRH	176
Chennai	SRH	KKR	SRH	KKR	114

Typical questions

- Which team played as Team 1 at maximum number of venues?
  - Count venues for Team 1 and take the max
    - One counter per team — but we don't know the teams or venues in advance!
  - Maintain a function, mapping teams to venues
  - A collection of (key,value) pairs — called a dictionary

City	Team 1	Team 2	Toss winner	Match winner	Run target
Chennai	RCB	CSK	RCB	CSK	174
Mohali	DC	PK	PK	PK	175
Kolkata	KKR	SRH	SRH	KKR	209
Jaipur	RR	LSG	RR	RR	194
Ahmedabad	GT	MI	MI	GT	169
Bengaluru	PK	RCB	RCB	RCB	177
Chennai	CSK	GT	GT	CSK	207
Hyderabad	SRH	MI	MI	SRH	278
Jaipur	RR	DC	DC	RR	186
Bengaluru	RCB	KKR	KKR	KKR	183
Lucknow	LSG	PK	LSG	LSG	200
Ahmedabad	SRH	GT	SRH	GT	163
Visakhapatnam	DC	CSK	DC	DC	192
Mumbai	MI	RR	RR	RR	126
Bengaluru	LSG	RCB	RCB	LSG	182
Visakhapatnam	KKR	DC	KKR	KKR	273
Ahmedabad	GT	PK	PK	PK	200
Hyderabad	CSK	SRH	SRH	SRH	166
Jaipur	RCB	RR	RR	RR	184
Mumbai	MI	DC	DC	MI	235
Lucknow	LSG	GT	LSG	LSG	164
Chennai	KKR	CSK	CSK	CSK	138
Mohali	SRH	PK	PK	SRH	183
Jaipur	RR	GT	GT	GT	197
Mumbai	RCB	MI	MI	MI	197
...	...	...	...	...	...
Chennai	SRH	RR	RR	SRH	176
Chennai	SRH	KKR	SRH	KKR	114

Typical questions

- Is winning the toss an advantage?

City	Team 1	Team 2	Toss winner	Match winner	Run target
Chennai	RCB	CSK	RCB	CSK	174
Mohali	DC	PK	PK	PK	175
Kolkata	KKR	SRH	SRH	KKR	209
Jaipur	RR	LSG	RR	RR	194
Ahmedabad	GT	MI	MI	GT	169
Bengaluru	PK	RCB	RCB	RCB	177
Chennai	CSK	GT	GT	CSK	207
Hyderabad	SRH	MI	MI	SRH	278
Jaipur	RR	DC	DC	RR	186
Bengaluru	RCB	KKR	KKR	KKR	183
Lucknow	LSG	PK	LSG	LSG	200
Ahmedabad	SRH	GT	SRH	GT	163
Visakhapatnam	DC	CSK	DC	DC	192
Mumbai	MI	RR	RR	RR	126
Bengaluru	LSG	RCB	RCB	LSG	182
Visakhapatnam	KKR	DC	KKR	KKR	273
Ahmedabad	GT	PK	PK	PK	200
Hyderabad	CSK	SRH	SRH	SRH	166
Jaipur	RCB	RR	RR	RR	184
Mumbai	MI	DC	DC	MI	235
Lucknow	LSG	GT	LSG	LSG	164
Chennai	KKR	CSK	CSK	CSK	138
Mohali	SRH	PK	PK	SRH	183
Jaipur	RR	GT	GT	GT	197
Mumbai	RCB	MI	MI	MI	197
...	...	...	...	...	...
Chennai	SRH	RR	RR	SRH	176
Chennai	SRH	KKR	SRH	KKR	114

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