

Lecture 1, 4 August 2026

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 — World Cup 2026

City	HTeam	ATeam	HGoals	Agoals	H XG	A XG
Mexico City	MEX	RSA	2	0	1.84	0.52
Zapopan	KOR	CZE	2	1	1.45	1.12
Toronto	CAN	BIH	1	1	1.35	0.98
Inglewood	USA	PAR	4	1	2.76	0.88
Foxborough	QAT	SUI	1	1	0.78	1.54
East Rutherford	BRA	MAR	1	1	1.62	1.1
Philadelphia	HAI	SCO	0	1	1.25	0.65
Seattle	AUS	TUR	2	0	1.48	0.72
Atlanta	GER	CUW	7	1	4.82	0.44
Santa Clara	NED	JPN	2	2	1.95	1.82
Houston	CIV	ECU	1	0	1.1	0.95
Kansas City	SWE	TUN	5	1	3.12	0.78
Miami Gardens	ESP	CPV	0	0	2.15	0.35
Arlington	BEL	EGY	1	1	1.58	1.12
Guadalupe	KSA	URU	1	1	0.85	1.76
...	...	...	...	...	...	...
Miami Gardens	FRA	ENG	4	6	2.88	2.88
East Rutherford	ESP	ARG	1	0	0.52	0.09

Questions

- How many matches were played?
- What was the maximum number of goals scored in a match?
- What was the average number of goals scored in a match?
- How many matches had above average number of goals?
- How many cities were venues?
- Which team played as Home Team at maximum number of venues?
- Is being Home Team an advantage?

Typical questions

- How many matches were played?
 - How is our table made available?
 - A **list** of rows, each a **tuple** of columns
 - `[row-1,...,row-N]`
 - `row-j` is `(City, HTeam, ATeam, HGoals, AGoals, HXG, AXG)`
 - 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

City	HTeam	ATeam	HGoals	AGoals	H XG	A XG
Mexico City	MEX	RSA	2	0	1.84	0.52
Zapopan	KOR	CZE	2	1	1.45	1.12
Toronto	CAN	BIH	1	1	1.35	0.98
Inglewood	USA	PAR	4	1	2.76	0.88
Foxborough	QAT	SUI	1	1	0.78	1.54
East Rutherford	BRA	MAR	1	1	1.62	1.1
Philadelphia	HAI	SCO	0	1	1.25	0.65
Seattle	AUS	TUR	2	0	1.48	0.72
Atlanta	GER	CUW	7	1	4.82	0.44
Santa Clara	NED	JPN	2	2	1.95	1.82
Houston	CIV	ECU	1	0	1.1	0.95
Kansas City	SWE	TUN	5	1	3.12	0.78
Miami Gardens	ESP	CPV	0	0	2.15	0.35
Arlington	BEL	EGY	1	1	1.58	1.12
Guadalupe	KSA	URU	1	1	0.85	1.76
...	...	...	...	...	...	...
Mexico City	MEX	ENG	2	3	1.88	1.61
Miami Gardens	POR	ESP	0	1	0.63	1.69
Santa Clara	USA	BEL	1	4	0.67	2.15
Kansas City	ARG	EGY	3	2	2.12	1.54
Seattle	SUI	COL	0	0	1.15	1.46
Foxborough	FRA	MAR	2	0	3.1	0.13
Inglewood	ESP	BEL	2	1	2.08	0.37
Miami Gardens	NOR	ENG	1	2	0.77	0.96
Kansas City	ARG	SUI	3	1	2.0	0.53
Arlington	FRA	ESP	0	2	0.3	1.63
Atlanta	ENG	ARG	1	2	0.54	1.80
Miami Gardens	FRA	ENG	4	6	2.88	2.88
East Rutherford	ESP	ARG	1	0	0.52	0.09

Typical questions

- What was the maximum number of goals scored in a match?
 - Again iterate through the rows
 - Maintain a variable `max` — the maximum goals seen so far
 - Initialize to 0 — lower bound, goals cannot be negative
 - Can also initialize `max` to goals in first row — maximum is not meaningful for an empty table
 - For each row, if goals exceeds `max`, update `max` to the current goals
 - At the end of the iteration, `max` is the largest number of goals scored

City	HTeam	ATeam	HGoals	Agoals	H XG	A XG
Mexico City	MEX	RSA	2	0	1.84	0.52
Zapopan	KOR	CZE	2	1	1.45	1.12
Toronto	CAN	BIH	1	1	1.35	0.98
Inglewood	USA	PAR	4	1	2.76	0.88
Foxborough	QAT	SUI	1	1	0.78	1.54
East Rutherford	BRA	MAR	1	1	1.62	1.1
Philadelphia	HAI	SCO	0	1	1.25	0.65
Seattle	AUS	TUR	2	0	1.48	0.72
Atlanta	GER	CUW	7	1	4.82	0.44
Santa Clara	NED	JPN	2	2	1.95	1.82
Houston	CIV	ECU	1	0	1.1	0.95
Kansas City	SWE	TUN	5	1	3.12	0.78
Miami Gardens	ESP	CPV	0	0	2.15	0.35
Arlington	BEL	EGY	1	1	1.58	1.12
Guadalupe	KSA	URU	1	1	0.85	1.76
...	...	...	...	...	...	...
Mexico City	MEX	ENG	2	3	1.88	1.61
Miami Gardens	POR	ESP	0	1	0.63	1.69
Santa Clara	USA	BEL	1	4	0.67	2.15
Kansas City	ARG	EGY	3	2	2.12	1.54
Seattle	SUI	COL	0	0	1.15	1.46
Foxborough	FRA	MAR	2	0	3.1	0.13
Inglewood	ESP	BEL	2	1	2.08	0.37
Miami Gardens	NOR	ENG	1	2	0.77	0.96
Kansas City	ARG	SUI	3	1	2.0	0.53
Arlington	FRA	ESP	0	2	0.3	1.63
Atlanta	ENG	ARG	1	2	0.54	1.80
Miami Gardens	FRA	ENG	4	6	2.88	2.88
East Rutherford	ESP	ARG	1	0	0.52	0.09

Typical questions

- What was the average number of goals scored in a match?
 - Need overall count and total sum
 - Know how to iterate and count
 - Total sum: iterate over rows and update variable `sum`
 - Initialize to 0
 - For each row, add current goals to `sum`
 - Average is `sum/count`
 - Naively, two iterations, one for `count` and another for `sum`
 - Single iteration, update `count` and `sum` with each row

City	HTeam	ATeam	HGoals	Agoals	H XG	A XG
Mexico City	MEX	RSA	2	0	1.84	0.52
Zapopan	KOR	CZE	2	1	1.45	1.12
Toronto	CAN	BIH	1	1	1.35	0.98
Inglewood	USA	PAR	4	1	2.76	0.88
Foxborough	QAT	SUI	1	1	0.78	1.54
East Rutherford	BRA	MAR	1	1	1.62	1.1
Philadelphia	HAI	SCO	0	1	1.25	0.65
Seattle	AUS	TUR	2	0	1.48	0.72
Atlanta	GER	CUW	7	1	4.82	0.44
Santa Clara	NED	JPN	2	2	1.95	1.82
Houston	CIV	ECU	1	0	1.1	0.95
Kansas City	SWE	TUN	5	1	3.12	0.78
Miami Gardens	ESP	CPV	0	0	2.15	0.35
Arlington	BEL	EGY	1	1	1.58	1.12
Guadalupe	KSA	URU	1	1	0.85	1.76
...	...	...	...	...	...	...
Mexico City	MEX	ENG	2	3	1.88	1.61
Miami Gardens	POR	ESP	0	1	0.63	1.69
Santa Clara	USA	BEL	1	4	0.67	2.15
Kansas City	ARG	EGY	3	2	2.12	1.54
Seattle	SUI	COL	0	0	1.15	1.46
Foxborough	FRA	MAR	2	0	3.1	0.13
Inglewood	ESP	BEL	2	1	2.08	0.37
Miami Gardens	NOR	ENG	1	2	0.77	0.96
Kansas City	ARG	SUI	3	1	2.0	0.53
Arlington	FRA	ESP	0	2	0.3	1.63
Atlanta	ENG	ARG	1	2	0.54	1.80
Miami Gardens	FRA	ENG	4	6	2.88	2.88
East Rutherford	ESP	ARG	1	0	0.52	0.09

Typical questions

- How many matches had above average number of goals?
 - 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 goals is above the average
 - **Filtered** update

City	HTeam	ATeam	HGoals	Agoals	H XG	A XG
Mexico City	MEX	RSA	2	0	1.84	0.52
Zapopan	KOR	CZE	2	1	1.45	1.12
Toronto	CAN	BIH	1	1	1.35	0.98
Inglewood	USA	PAR	4	1	2.76	0.88
Foxborough	QAT	SUI	1	1	0.78	1.54
East Rutherford	BRA	MAR	1	1	1.62	1.1
Philadelphia	HAI	SCO	0	1	1.25	0.65
Seattle	AUS	TUR	2	0	1.48	0.72
Atlanta	GER	CUW	7	1	4.82	0.44
Santa Clara	NED	JPN	2	2	1.95	1.82
Houston	CIV	ECU	1	0	1.1	0.95
Kansas City	SWE	TUN	5	1	3.12	0.78
Miami Gardens	ESP	CPV	0	0	2.15	0.35
Arlington	BEL	EGY	1	1	1.58	1.12
Guadalupe	KSA	URU	1	1	0.85	1.76
...	...	...	...	...	...	...
Mexico City	MEX	ENG	2	3	1.88	1.61
Miami Gardens	POR	ESP	0	1	0.63	1.69
Santa Clara	USA	BEL	1	4	0.67	2.15
Kansas City	ARG	EGY	3	2	2.12	1.54
Seattle	SUI	COL	0	0	1.15	1.46
Foxborough	FRA	MAR	2	0	3.1	0.13
Inglewood	ESP	BEL	2	1	2.08	0.37
Miami Gardens	NOR	ENG	1	2	0.77	0.96
Kansas City	ARG	SUI	3	1	2.0	0.53
Arlington	FRA	ESP	0	2	0.3	1.63
Atlanta	ENG	ARG	1	2	0.54	1.80
Miami Gardens	FRA	ENG	4	6	2.88	2.88
East Rutherford	ESP	ARG	1	0	0.52	0.09

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	HTeam	ATeam	HGoals	Agoals	H XG	A XG
Mexico City	MEX	RSA	2	0	1.84	0.52
Zapopan	KOR	CZE	2	1	1.45	1.12
Toronto	CAN	BIH	1	1	1.35	0.98
Inglewood	USA	PAR	4	1	2.76	0.88
Foxborough	QAT	SUI	1	1	0.78	1.54
East Rutherford	BRA	MAR	1	1	1.62	1.1
Philadelphia	HAI	SCO	0	1	1.25	0.65
Seattle	AUS	TUR	2	0	1.48	0.72
Atlanta	GER	CUW	7	1	4.82	0.44
Santa Clara	NED	JPN	2	2	1.95	1.82
Houston	CIV	ECU	1	0	1.1	0.95
Kansas City	SWE	TUN	5	1	3.12	0.78
Miami Gardens	ESP	CPV	0	0	2.15	0.35
Arlington	BEL	EGY	1	1	1.58	1.12
Guadalupe	KSA	URU	1	1	0.85	1.76
...	...	...	...	...	...	...
Mexico City	MEX	ENG	2	3	1.88	1.61
Miami Gardens	POR	ESP	0	1	0.63	1.69
Santa Clara	USA	BEL	1	4	0.67	2.15
Kansas City	ARG	EGY	3	2	2.12	1.54
Seattle	SUI	COL	0	0	1.15	1.46
Foxborough	FRA	MAR	2	0	3.1	0.13
Inglewood	ESP	BEL	2	1	2.08	0.37
Miami Gardens	NOR	ENG	1	2	0.77	0.96
Kansas City	ARG	SUI	3	1	2.0	0.53
Arlington	FRA	ESP	0	2	0.3	1.63
Atlanta	ENG	ARG	1	2	0.54	1.80
Miami Gardens	FRA	ENG	4	6	2.88	2.88
East Rutherford	ESP	ARG	1	0	0.52	0.09

Typical questions

- Which team played as Home Team at maximum number of venues?
 - Count venues for Home Team 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	HTeam	ATeam	HGoals	Agoals	H XG	A XG
Mexico City	MEX	RSA	2	0	1.84	0.52
Zapopan	KOR	CZE	2	1	1.45	1.12
Toronto	CAN	BIH	1	1	1.35	0.98
Inglewood	USA	PAR	4	1	2.76	0.88
Foxborough	QAT	SUI	1	1	0.78	1.54
East Rutherford	BRA	MAR	1	1	1.62	1.1
Philadelphia	HAI	SCO	0	1	1.25	0.65
Seattle	AUS	TUR	2	0	1.48	0.72
Atlanta	GER	CUW	7	1	4.82	0.44
Santa Clara	NED	JPN	2	2	1.95	1.82
Houston	CIV	ECU	1	0	1.1	0.95
Kansas City	SWE	TUN	5	1	3.12	0.78
Miami Gardens	ESP	CPV	0	0	2.15	0.35
Arlington	BEL	EGY	1	1	1.58	1.12
Guadalupe	KSA	URU	1	1	0.85	1.76
...	...	...	...	...	...	...
Mexico City	MEX	ENG	2	3	1.88	1.61
Miami Gardens	POR	ESP	0	1	0.63	1.69
Santa Clara	USA	BEL	1	4	0.67	2.15
Kansas City	ARG	EGY	3	2	2.12	1.54
Seattle	SUI	COL	0	0	1.15	1.46
Foxborough	FRA	MAR	2	0	3.1	0.13
Inglewood	ESP	BEL	2	1	2.08	0.37
Miami Gardens	NOR	ENG	1	2	0.77	0.96
Kansas City	ARG	SUI	3	1	2.0	0.53
Arlington	FRA	ESP	0	2	0.3	1.63
Atlanta	ENG	ARG	1	2	0.54	1.80
Miami Gardens	FRA	ENG	4	6	2.88	2.88
East Rutherford	ESP	ARG	1	0	0.52	0.09

Typical questions

- Is being Home Team an advantage?

City	HTeam	ATeam	HGoals	Agoals	H XG	A XG
Mexico City	MEX	RSA	2	0	1.84	0.52
Zapopan	KOR	CZE	2	1	1.45	1.12
Toronto	CAN	BIH	1	1	1.35	0.98
Inglewood	USA	PAR	4	1	2.76	0.88
Foxborough	QAT	SUI	1	1	0.78	1.54
East Rutherford	BRA	MAR	1	1	1.62	1.1
Philadelphia	HAI	SCO	0	1	1.25	0.65
Seattle	AUS	TUR	2	0	1.48	0.72
Atlanta	GER	CUW	7	1	4.82	0.44
Santa Clara	NED	JPN	2	2	1.95	1.82
Houston	CIV	ECU	1	0	1.1	0.95
Kansas City	SWE	TUN	5	1	3.12	0.78
Miami Gardens	ESP	CPV	0	0	2.15	0.35
Arlington	BEL	EGY	1	1	1.58	1.12
Guadalupe	KSA	URU	1	1	0.85	1.76
...	...	...	...	...	...	...
Mexico City	MEX	ENG	2	3	1.88	1.61
Miami Gardens	POR	ESP	0	1	0.63	1.69
Santa Clara	USA	BEL	1	4	0.67	2.15
Kansas City	ARG	EGY	3	2	2.12	1.54
Seattle	SUI	COL	0	0	1.15	1.46
Foxborough	FRA	MAR	2	0	3.1	0.13
Inglewood	ESP	BEL	2	1	2.08	0.37
Miami Gardens	NOR	ENG	1	2	0.77	0.96
Kansas City	ARG	SUI	3	1	2.0	0.53
Arlington	FRA	ESP	0	2	0.3	1.63
Atlanta	ENG	ARG	1	2	0.54	1.80
Miami Gardens	FRA	ENG	4	6	2.88	2.88
East Rutherford	ESP	ARG	1	0	0.52	0.09

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