

Introduction to Data Science Using Python

9 – 13 September 2024

Proudly brought to you by the School for Data Science and Computational Thinking at Stellenbosch University, Stellenbosch, South Africa.

1. Overview:

More data is being generated and stored than ever before. We generate data through various processes. For example, we generate data when we transact at a shop, when our smart watches record our steps, and when we record a Zoom meeting. Given the advancement in computational power and storage, we are using these stored datasets to make sense of the world around us and to make decisions. These decisions can help to advance business, health, sporting and other outcomes. An important aspect of learning data science is learning to code and using code to build data science models. This workshop introduces data science using Python. It builds up from basic coding in Python to data science models.

2. Outcomes:

Participants will gain competency in the following:

- Fundamental Python programming concepts
- Data manipulation and analysis with Python
- Introduction to data visualization
- Predictive analytics
- Basic Machine Learning
- Python libraries such as Pandas, Matplotlib, Scikit Learn, Seaborn, Statsmodels, Numpy
- Practical exercises and real-world examples

3. Syllabus and Schedule:

The workshop is scheduled to take place from September 9th to September 13th, 2024, with sessions held daily from 9:00 AM to 3:00 PM South African Standard Time (SAST).

The topic schedule is shown below:

Mon. 9 Sept 2024		Introduction, Cause and Effect 9:00 - 10:30 SAST	Tables and Data Types 10:40- 11:40 SAST	Groups, Pivots, and Joins 11:50 - 12:50 SAST	Visualization- Charts, Histogram 13:00-13:30 SAST
Tues. 10 Sept 2024		Flowcharts 9:00 - 10:00 SAST	Conditionals and Iterations 10:10 - 11:10 SAST	Functions 11:20 - 12:20 SAST	Chance 12:30 - 13:30 SAST
Wed. 11 Sept 2024		Sampling 9:00-10:00 SAST	Models 10:10-11:10 SAST	Comparing Distributions, Decision and Uncertainty 11:20 - 12:20 SAST	A/B Testing 12:30-13:30 SAST
Thurs. 12 Sept 2024		Causality 9:00-10:00 SAST	Confidence and Intervals Interpreting Confidence 10:10 - 11:10 SAST	Center and Spread The Normal Distribution 11:20-12:20 SAST	Correlation 12:30-13:30 SAST
Fri. 13 Sept 2024		Linear Regression 9:00-10:00 SAST	Regression Inference 10:10-11:10 SAST	Classification 11:20-12:20 SAST	CLOSING REMARKS

4. Presenters will include:

- Dr. Sunday Oladejo
- Dr. Marike Visser
- Prof. Kanshukan Rajaratnam

5. Other pertinent information:

This workshop will be conducted online, and weblinks for the sessions will be provided to all registered participants.

A detailed workshop schedule, datasets, and cheat sheets will be sent out well in advance to ensure everyone is prepared for the event.

All participants are required to attend all sessions, undertake some homework between sessions, and participate in all sessions. Participants must attend all sessions because each session builds up on prior work. By participating, you agree for the workshop to be recorded. The workshop is recorded for School purposes and will not be shared with the participants.

Please note, this is not a short course or a module. Academic credits will not be granted for completing or attending this workshop and certificates will not be issued. We will also not be writing letters of completion for attending this workshop.