# Regional Training on Data Science for Official Statistics

**11-15 May, 2026**

Daejeon, Republic of Korea

## I. Background

In recent years, the data landscape has evolved rapidly, creating for National Statistical Offices (NSOs) both significant opportunities for innovation and important challenges in maintaining data quality and reliability. In response to these developments, the Statistical Institute for Asia and the Pacific (SIAP) and the Ministry of Data And Statistics (MODS) are organizing a Regional Training on Data Science for Official Statistics, recognizing both the opportunities and the challenges associated with integrating Data Science methods and tools into NSO processes.

The overall objective of the week-long training is to build the capacity of NSOs to integrate Data Science methods into the production of Official Statistics. It seeks to strengthen participants' skills in handling, analyzing, and visualizing data sources, with a particular focus on reproducible methods and tools for monitoring and reporting Official Statistics and Sustainable Development Goals (SDGs) indicators.

Through a combination of lectures, case studies, and practical exercises, the course will demonstrate how Data Science, Data Visualization, advanced analytics, and reproducible methods can be effectively integrated into established statistical practices within NSOs. The training will equip participants with the skills and knowledge needed to produce trustworthy statistics, graphics, and analytical outputs.

The overall rationale is to enable participants to adapt to an evolving data ecosystem characterized by new data sources, changing user demands, and emerging analytical approaches, while sustaining and reinforcing the relevance and trustworthiness of Official Statistics. In this context, the use of open-source software is integral to the training approach and allow participants to apply the acquired skills beyond the training, thereby fostering sustainable capacity development within NSOs.

This investment in human capital is expected to contribute to the production of high-quality and reliable outputs, thereby supporting more informed decision-making, effective public policies, and accelerated progress in the production and use of Official Statistics and SDG indicators.

## II. Target Audience

The course is designed for mid-level statisticians and technical staff from National Statistical Offices in Asia and the Pacific, whose main responsibilities include the production and dissemination of Official Statistics and SDG indicators. Approximately 16 participants will be invited to attend. The participants are expected to have a basic proficiency in the analysis and modelling of statistical data and experience or appetite in the use of a modern statistical software or language (R, Python).

- Experience handling various data sources in their professional work
- Prior experience using R, Python, or other statistical software
- Likelihood of working on Data applications and in the computation of Official Statistics

## III. Course design and content

The course is comprised of several modules, consisting of presentations, exercises, and handouts on methods and topics used in Data Science, including illustrative good practices from resource persons presenting practical approaches, modern processes and solutions to commonly encountered issues; sharing of relevant country experiences as well as handouts sessions with practical exercises. Plenary and group discussions will be organized throughout the course. Some prior experience and knowledge in programming would also help participants understand some examples and case studies.

The topics covered in this course include:

- Data Science methods and tools
- Data Visualization good practices
- Reproducible analytical pipelines and data science good practices
- Introduction to advanced Data Science methods (Machine Learning, AI)
- Biases and ethics in Data Science
- Challenges of Data Science integration

## IV. How do I prepare for this training?

- Prior to the training, participants will be required to **complete** **a preliminary tutorial** on the use of R. This self-paced online tool will provide instructions on how to install R and RStudio as well as recall a few elements of the R programming language.
- Additionally, to facilitate exchange of experiences, **each participant will need to submit a country presentation** of no more than 3 slides on or **before Wednesday 6 May, 2026** to escap-siap@un.org and christophe.bontemps@un.org. Participants will be required to deliver their country presentations on day one of the training. Each participant will be allocated a maximum of 5 minutes for their oral presentation. It should briefly introduce the process in the creation of Official Statistics, present main challenges encountered and outline specific plans to improve statistical production processes. More information will be provided prior to the course.

## V. Learning Objectives

At the end of the course participants should be able to:

- **Explain** where Data Science fits within NSO statistical production workflows
- **Explore** Data Science tools and methods
- **Create** clear, reproducible statistics indicators and graphics for statistical communication
- **Apply** reproducible workflows using code, documentation, and versioned outputs
- **Select** and **use** appropriate Data Science tools for common NSO tasks
- **Combine** and **analyze** multiple data sources for simple analytical use cases
- **Experiment** with selected advanced Data Science methods in existing workflows
- **Discuss** opportunities and limitations of AI for Official Statistics

## VI. Dates and venue

The workshop will take place over five days, from **11-15 May**, in Daejeon.

Participants are required to bring their own laptop equipped with the latest version of R/RStudio (see the instructions in the mandatory tutorial )