SDMX EXPERTS WORKSHOP ADVANCED CAPACITY BUILDING

Low code SDMX data pipelines automation

9.00-12.30, Friday 11 October 2024

PARK CENTRAAL AMSTERDAM Stadhouderskade 25 1071 ZD Amsterdam

What will be covered?

This advanced capacity building event is designed to empower participants with the knowledge and skills in SDMX data processing using Python, leveraging AI with SDMX endpoints, and strategies for adopting advanced metadata management practices.

Who should join?

Data practitioners (statisticians, data engineers, data scientists, data analysts, data managers) already familiar with SDMX and data management practices.

Introduction (15 minutes)

- Session objectives and learning outcomes.
- Introduction to the importance of SDMX data pipelines automation and 'low code'[1] paradigms and their impact on SDMX data management.

Module 1: SDMX data processing with Python (1 hour)

- Learn how to set up your Python environment for SDMX manipulation.
- Explore the available libraries and tools incl. pysdmx, dotstat-io, mappings-utility.
- Fetch and process data from SDMX APIs using pysdmx.

Module 2: Elevate Python with AI, LLMs, and chatbots (1 hour)

- Learn how to leverage Large Language Models (LLMs) in Python, with a focus on increasing complexity to enhance performance.
- Explore how to store segmented documents in a vector database and perform effective semantic searches.
- Develop a chatbot that interacts with SDMX metadata, rather than relying on the language model's built-in knowledge.

Note: This session will be immersive with a high level of interaction and hands-on activities.

Module 3: Mastering advanced metadata management (1 hour)

- This module features three introductory presentations that delve into:
  - The strategic importance of metadata in automating SDMX data pipelines,
  - Leveraging metadata for dataset interactions and integration,
  - Crafting intelligent, self-descriptive data systems with SDMX.
- **Discussion:** Challenges and strategies for implementing advanced metadata solutions.

[1] **Low-code** basically means that you need less lines of code compared to "classic" coding, which makes the code easier to write, easier to understand and easier to maintain. For example, *Low-code Python* allows users to use one of the most popular scripting languages in the world — Python — with all the benefits of the low-code paradigm. Low-code Python saves time and allows users with a less technical background to implement the solutions that they need, faster!

Wrap-up (15 minutes)

- Final Q&A.
- Recap of key takeaways, tools/resources for further Learning.

Prerequisites: Familiarity with Python, basic understanding of SDMX, and experience with data management practices are assumed for this session. Participants should have a laptop (computer) with Python installed and access to the internet to access a relevant SDMX APIs. *Please note that there is an expectation that participants are not beginners as the hands-on exercises will be more a following of steps as opposed to intense instructor engagement to resolve issues.*