

# ► Labour Market Intelligence: Drafting Analytical Reports from SDMX Data Using Generative AI

Shutong Ding

Department of Statistics, ILO

29/09/2025

## ► Labour Market Information System (LMIS)

### ILO supports countries in developing robust LMIS

A LMIS is a network of **institutions**, **persons** and **information** that have mutually recognized roles, agreements and functions with respect to the production, **storage**, dissemination and use of labour market related data and metadata.

- Ensure timely and harmonized labour market data
- Data dissemination using the open-source **.Stat Suite** platform (**SDMX compliant**)
- Nearly 30 LMIS projects currently impacting over 50 countries
- Increasing volume of publicly available data presents both opportunities and **analysis challenges**



## Examples of a LMIS Data Explorer

### Time-Related Underemployment (TRU) by sex, strata and status in employment

The proportion of employed persons who worked less than a specified threshold of hours during the reference period in all their jobs and who were available and willing to work more hours. This indicator highlights the extent of underutilisation of labour among employed individuals.

**Dimensions:** Sex, Strata, Status in Employment, Time Period

[Download](#)

### Unemployment Rate by sex and educational level

It is the proportion of the unemployed population to the Total Labour Force (Employed + Unemployed). It reflects the inability of an economy to generate employment for those who are available and actively seeking work. It is therefore seen as an indicator of the efficiency and effectiveness of an economy and the performance of the labour market.

**Dimensions:** Educational Level, Sex, Time Period

[Download](#)

### Unemployment Rate by sex and strata

It is the proportion of the unemployed population to the Total Labour Force (Employed + Unemployed). It reflects the inability of an economy to generate employment for those who are available and actively seeking work. It is therefore seen as an indicator of the efficiency and effectiveness of an economy and the performance of the labour market.

**Dimensions:** Sex, Strata, Time Period

[Download](#)

### Unemployment Rate by sex, strata and disability status

It is the proportion of the unemployed population to the Total Labour Force (Employed + Unemployed). It reflects the inability of an economy to generate employment for those who are available and actively seeking work. It is therefore seen as an indicator of the efficiency and effectiveness of an economy and the performance of the labour market.

**Dimensions:** Disability Status, Strata, Sex, Time Period

[Download](#)

### Unemployment by sex and strata

The unemployed population are individuals who are not in employment, seeking work and are available to start working if a job opportunity arises.

**Dimensions:** Sex, Strata, Time Period

[Download](#)

### Developer API query builder

The application programming interface (API) based on the SDMX standard allows a developer to programmatically access the data using simple RESTful URL and HTTP header options for various choices of response formats including JSON. To get started check the [API documentation](#). For any question [contact us](#).

#### Data query

SDMX flavour: [Flat](#) [Time series](#)

```
https://sdmx.lmis.hrdc.org.bw/rest/data/HRDC,DF_NEET_SEX_REF_AREA,1.0/all?dimensionAtObservation=AllDimensions
```

[Copy code](#)

#### Structure query

```
https://sdmx.lmis.hrdc.org.bw/rest/dataflow/HRDC/DF_NEET_SEX_REF_AREA/1.0?references=all
```

[Copy code](#)

The query filter is generated according to the current data selection. To change the data selection, use the filters on the left.

### Youth Not in Employment Education or Training (NEET) (SDG 8.6.1) by sex and census district i

Frequency: Quarterly • Indicator: Youth Not in Employment Education or Training (NEET) (SDG 8.6.1)

Combined unit of measure: Rate, Units

	Time Period	2019-Q4	2020-Q4	2021-Q4	2022-Q4	2023-Q3
Census District						
Sex: Total						
Total		36.1	37.5	39.5	38.5	39.9
GABORONE CITY		27.9	25.3	23.1	25.5	27.0
FRANCISTOWN CITY		29.8	26.6	33.6	35.5	35.9
LOBATSE		34.1	25.6	26.9	39.3	32.5
SELIBE PHIKWE		40.3	37.3	45.5	31.5	29.3
ORAPA		38.5	17.0	42.5	32.8	41.6
JWANENG		24.9	16.7	15.5	17.7	47.1
SOWA		27.1	14.6	16.7	24.0	38.5
NGWAKETSE / SOUTHERN		53.8	50.5	54.2	42.4	44.9
BAROLONG		50.2	41.8	50.9	42.5	40.7
NGWAKETSE WEST		46.9	37.3	39.7	48.8	45.9
SOUTH EAST		22.6	23.4	28.8	24.9	31.5

Source: Labour Market Observatory, Botswana

## ► Feasibility check

### Data

- LMIS database
- Aggregated data, no micro-level disclosure
- Well structured and harmonized
- SDMX handles both data and metadata via API
- Indicator typically with demographic breakdowns

### AI tools

- Any existing AI tools to achieve this by direct prompts?  
**Not yet fully achievable**
- Powerful LLMs, paid access via API: OpenAI
- Require dedicated AI-driven workflow and structured prompting

### Desired solution

- Affordable
- “Zero-hallucination”
- Meaningful insights
- Draft - not final, not replacing
- Minimized configuration
- Centralized deployment & LLM management preferred
- User login with credit allocation
- Python-based web application

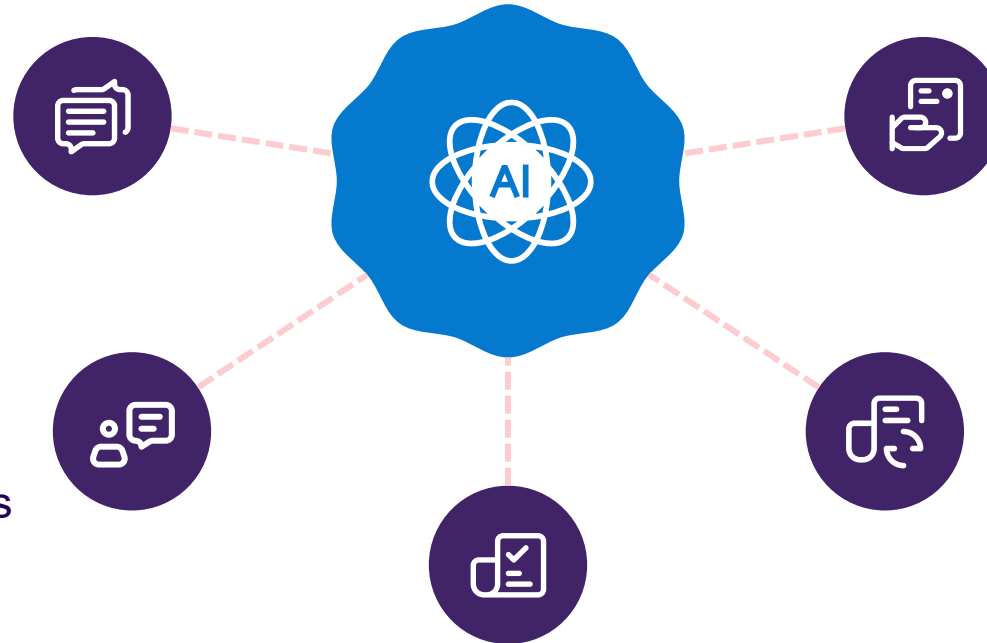
# GENERATIVE AI

## Support **multilingual**

When handling data from different countries, language will not be an issue.

## **Learn** knowledge & report styles

Integrate expert knowledge and the report's tone, style, and interpretation methods into the model by selecting either a brief, summary, or full report format.



## **Draft** analytical reports

Produce analytical reports and offer insights based on the provided SDMX data.

## **Reproduce** effortlessly

The analysis of one table can be generalized to conduct similarly for others; however, exact same results **may not be guaranteed**.

## Create **code** for graph plotting

Allow the model to select the graph types and data slides for presentation by generating Python code to create the charts.



# Solution

## Labour Market Information System AI Report Assistant

Bringing Artificial Intelligence to LMIS

Instant Analysis AI-Powered SDMX API

BETA



### Single Table Analysis

#### Analysis in Progress

Started at: ...

- ✓ Step 1 of 7 - Downloading CSV data...
- ✓ Step 2 of 7 - Querying for dataflow info...
- ✓ Step 3 of 7 - Data cleaning...
- ✓ Step 4 of 7 - Generate the graph code and report
- ⌚ Step 5 of 7 - Generate the summary report ...

Please wait while your analysis is being processed. This may take a few moments.



### Overall Market Report

#### Introduction

This report presents a comprehensive analysis of the labour market in Botswana, utilizing data by Sex, Age Group, and Educational Level" (SDMX Code: HRDC:DF\_EMP\_SEX\_AGE\_EDU(1.0)). The employment trends and dynamics within the country, providing insights into how various demographic factors influence the labour market. The data is sourced from Statistics Botswana and is reported on a quarterly basis.

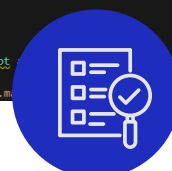
#### ## Scope and Data Summary

The dataset covers employment data from the fourth quarter of 2019 to the third quarter of 2023. The analysis is on the most recent period, 2023-Q3, while also considering trends from previous periods. Key variables in the dataset include employment figures disaggregated by gender, age group, and educational level, providing a clear picture of employment levels across different demographic groups.

#### # Data setup & loading

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

pd.set_option('display.m
```



### Dashboard Template



## Workflow & Methodology – Phase I

### Single table analysis

- ▶ **Data Retrieval:** API calls to SDMX entry point
- ▶ **Data processing:**
  - ▶ Reshape data for context
  - ▶ Slice historical data series and recent periods
- ▶ **AI analysis:** Call LLM via API to generate narrative reports and Python code for graphs.
- ▶ **Consolidation**

#### AI Insights

ⓘ Please note: this AI-generated content may not always be accurate or up-to-date.

### Statistical Report: Employment Trends in Botswana (2019-Q4 to 2023-Q3)

#### Introduction

This report presents a comprehensive analysis of Botswana's labour market using the dataset **"Employment by sex, age group and educational level"** (SDMX code: HRDC:DF\_EMP\_SEX\_AGE\_EDU(1.0)), sourced from Statistics Botswana. The dataset provides quarterly employment figures disaggregated by sex, age group, and educational attainment, offering critical insights into the structure and dynamics of Botswana's workforce. Such granular data is essential for policymakers, researchers, and stakeholders aiming to understand employment patterns, identify disparities, and design evidence-based interventions to foster inclusive and sustainable labour market growth.

#### Scope and Data Summary

The analysis covers the period from the fourth quarter of 2019 (2019-Q4) to the third quarter of 2023 (2023-Q3), with a particular focus on the most recent data available (2023-Q3). The dataset encompasses:

**Employment counts** (in thousands) for the total population, as well as breakdowns by sex (male, female), age groups (from 15-17 up to 75+), and educational levels (never attended, primary, secondary, vocational, tertiary, non-formal, and not applicable).

**Quarterly frequency**, enabling the identification of both short-term fluctuations and longer-term trends.

**National coverage**, with all census districts included.

#### Key Statistics

**Total employment** reached a peak of **788,616 persons in 2023-Q3**, the highest in the observed period.

**Lowest employment** was recorded in **2021-Q4** at **717,418 persons**, reflecting the impact of the COVID-19 pandemic.

**Male employment** in 2023-Q3: 40

**Dominant educational attainment**

**Largest age cohort employed:** 25

#### View & Edit Plot Code

```
9
10 # Data preparation (ensure chronological order)
11 df = df.sort_values("TIME_PERIOD")
12
13 # Create figure
14 fig = px.line(
15     df,
16     x="TIME_PERIOD",
17     y="OBS_VALUE",
18     markers=True,
19     title="Botswana Total Employment Trends (2019-Q4 to 2023-Q3)",
20     labels={
21         "TIME_PERIOD": "Quarter",
22         "OBS_VALUE": "Employment (Thousands of Persons)"
23     }
24 )
25
```

[Run Updated Code](#)[Reload Code](#)

## ► Workflow & Methodology – Phase II (Ongoing)

### Overall market report

- **Roll-out:** Execute single-table analysis for all tables.
- **Knowledge-base generation:**
  - Provide historical labour market report (e.g., LFS releases) to LLM
  - Derive insights on report layouts, tone, preferred charts, and causal analysis
- **Consolidation:** Combine individual report outputs and learned report templates to generate final report drafts in Jupyter Notebook.



## Workflow & Methodology – Phase II (Ongoing)

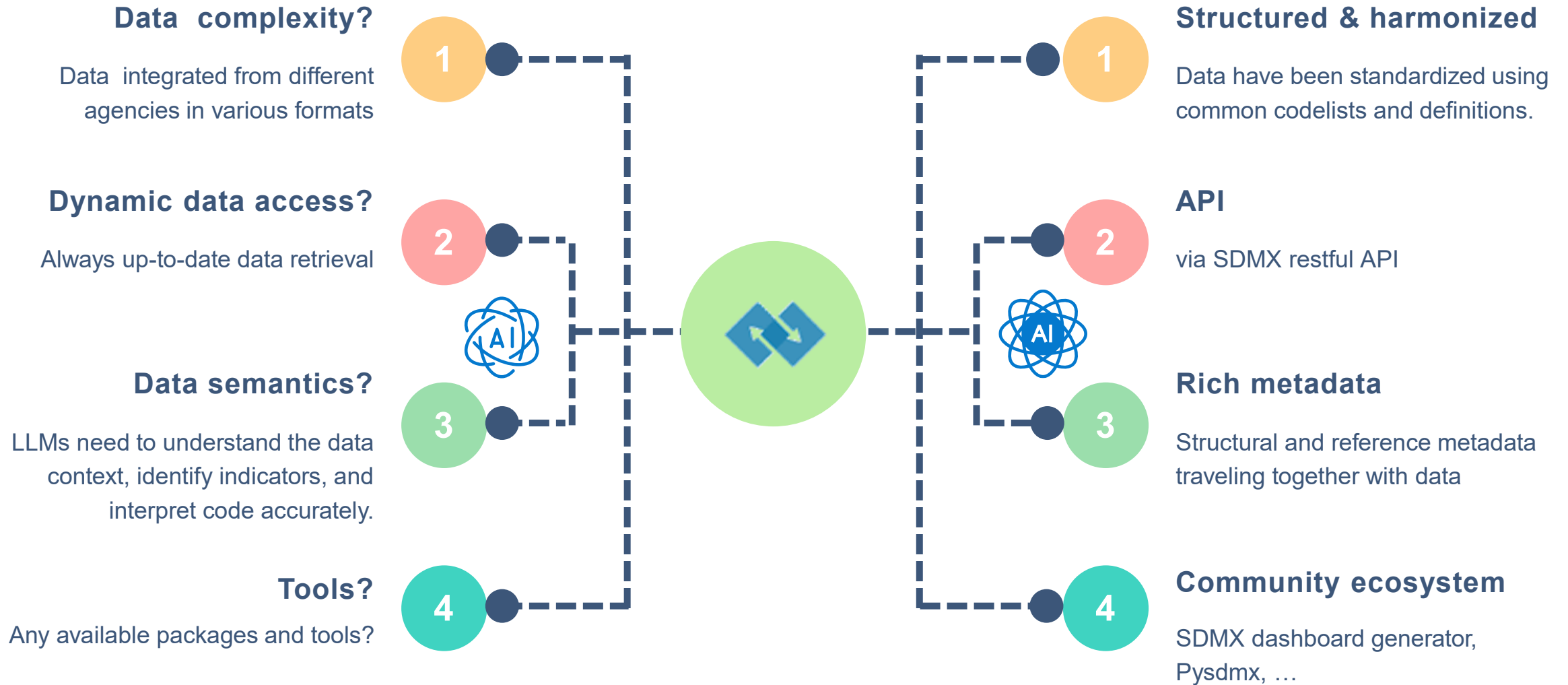
### Dashboard template

- ▶ LLM as in overall market report
- ▶ **Dashboard Generator:** Existing open-source tool using SDMX data query in JSON (2023 SDMX GC Hackathon Winner)
- ▶ **LLM Learning:** Consume existing dashboard JSON to learn about output template
- ▶ **Template generation**

```
"columns": [
  {
    "subtitle": {
      "align": "center"
    },
    "xAxisConcept": "NATION",
    "yAxisConcept": "OBS_VALUE",
    "data": "https://sdmx-jo.lmis.systems/rest/data/J0110,DF_AXJOW_NATION_SEX,1.0/.F.A.EGY+TUN+SYR+YEM+IRQ+SDN+M"
  }
],
{
  "columns": [
    {
      "id": "EMP_AREA_SEX",
      "type": "column",
      "colSize": 3,
      "title": {
        "text": {
          "en": "Employed by governorate",
          "ar": "*-*-*"
        },
        "size": "24px",
        "weight": "bold",
        "align": "center"
      },
      "subtitle": {
        "text": {
          "en": "Distribution by Sex - {TIME_PERIOD}",
          "ar": "*-*-*"
        },
        "size": "20px",
        "weight": "normal",
        "align": "center"
      },
      "legend": {
        "concept": "SEX",
        "location": "bottom"
      },
      "xAxisConcept": "REF_AREA",
      "data": "https://sdmx-jo.lmis.systems/rest/data/J0110,DF_EMP_AREA_SEX,1.0/34+33+32+31+22+23+24+21+13+12+11.A"
    }
  ],
  "columns": [
```

# SDMX Empowering AI

10



## ► Points for reflection

- **SDMX** naturally facilitates automation and AI integration.
- Simply offering products in multiple languages can be costly – AI makes it seamless.
- “Zero-hallucination” remains questioned, expected to have further insights from our field-testing.
- AI vs predetermined programming **trade-off**: More AI, richer insights but increased error risk (e.g., autogenerated Python code).
- Would it be possible to **generalize** the AI analytical core beyond the labour context?

# Thank you!

## Meet Our Team:

- Edgardo Greising
- Christophe Vittorelli
- Shutong Ding
- Weichen Lei
- Yunxiang Guo
- Iulian Pogor

 [Imis@ilo.org](mailto:Imis@ilo.org)  
[ding@ilo.org](mailto:ding@ilo.org)

