**Ministry of Education and Research**

**of the Republic of Moldova**

**Technical University of Moldova**

**Faculty of Computers, Informatics and Microelectronics**

**Discipline: Exploratory Data Analysis**

**REPORT**

**Assignment 1:** European Development Indicators

Performed by:

st. gr. IS-251-M Igor Latii

Verified by:

l. univ. Alexandru Monahov

Chişinău - 2025

**Assignment 1: European Development Indicators**

**Phase 1: Country Selection & Data Acquisition**

1. **Chosen Country**

For this project, the selected European country is Latvia (LV). Latvia represents an interesting case for economic analysis due to its relatively small and open economy, strong dependence on trade and transport flows, and demographic challenges such as migration and population decline. The country’s integration into the EU single market and its adoption of the euro have also made it highly sensitive to regional and global economic dynamics.

1. **Research Questions**

The following research questions (RQ) have been defined to explore the economic relationships and structural dynamics in Latvia using quantitative indicators from Eurostat and the World Bank:

**RQ1**: How has the evolution of external trade and passenger flows correlated with Latvia’s GDP and overall economic activity?

**RQ2**: What is the relationship between unemployment, migration, and international departures of Latvian residents?

**RQ3**: How do freight and passenger transport volumes relate to inflation?

These questions aim to go beyond simple descriptive analysis and investigate the interconnections between trade, mobility, and macroeconomic performance over time.

1. **Selected Indicators (Eurostat / World Bank)**

To address the above research questions, a set of 15 indicators was selected, primarily from Eurostat, with one Net Migration (SM.POP.NETM) sourced from the World Bank via the link: <https://data.worldbank.org/indicator/SM.POP.NETM>, as this measure is not available in Eurostat.

The final selection is documented in [/reports/indicators.csv](indicators.csv) and [/reports/indicators.xlsx](indicators.xlsx) and includes 6 monthly, 2 quarterly, 1 semestrial, and 6 annual datasets.

Mapping of Research Questions to Indicators can be presented as follows:

**RQ Indicators (Code)**

RQ1 namq\_10\_gdp, nama\_10\_exi, avia\_paocc, road\_pa\_mov, sts\_inpr\_m, sts\_trtu\_m, tour\_occ\_nim, nrg\_pc\_202

RQ2 une\_rt\_m, lfsi\_emp\_q, migr\_emi1ctz, SM.POP.NETM, demo\_pjan, tour\_occ\_nim

RQ3 prc\_hicp\_manr, tran\_hv\_frtra, avia\_paocc, road\_pa\_mov, sts\_inpr\_m, sts\_trtu\_m, nrg\_pc\_202

This selection ensures a balanced dataset across economic, demographic, and transport sectors, aligned with the research focus.

1. **Data Sources and Acquisition Approach**

Primary source: **Eurostat** — official statistical database of the European Union, accessed programmatically via the eurostat Python package.

Secondary source: **World Bank Open Data** — used exclusively for the Net Migration (SM.POP.NETM) indicator.

Data collection was automated through a custom Python script [/src/collecting\_data.py](../src/collecting_data.py), which:

1. Reads the list of indicators from [/reports/indicators.csv](indicators.csv).
2. Uses the Eurostat API to download each dataset.
3. Filters observations by the selected country (geo = LV).
4. Saves the resulting tables in [/data/raw/](../data/raw) for further processing.

Each dataset is stored in its raw format (CSV) to ensure full reproducibility and transparency of preprocessing.

The World Bank dataset for Net Migration was downloaded manually in CSV format from (<https://data.worldbank.org/indicator/SM.POP.NETM>).

1. **Data Frequency and Coverage**

The collected datasets vary in temporal frequency:

* Monthly indicators: Inflation, Unemployment rate, Industrial production, Retail trade turnover, Tourism nights spent, Air passenger transport.
* Quarterly indicators: GDP, Employment.
* Semestrial indicator: Energy consumption per capita.
* Annual indicators: Exports and imports, Migration, Freight, Road passenger transport, Population.

The temporal coverage spans from the 1950s to 2025, depending on data availability per series.

Specific data frequency and coverage of each selected indicator is documented in [/reports/indicators.xlsx](indicators.xlsx).

More than half of the indicators (8 out of 15) have monthly or quarterly frequency, satisfying the requirement for high-frequency time series suitable for short-term dynamics analysis.

All data were validated to include only Latvia (geo = LV), ensuring full consistency across datasets and comparability between indicators.

**Phase 2: Data Cleaning and Preparation**