

Drivers of economic growth in Europe: The role of education and Household savings (2013-2023)

- **Author:** Strahinja Miljkovic
- **Date:** February 2026
- **Tools used:** R, SQL(Google BigQuery) and Tableau
- **Data source:** Eurostat

Contents

1. Executive summary	3
2. Business problem	4
3. Data Sources and Preparation	5
3.1 Data sources	5
3.2 Data cleaning in R.....	6
4. Analysis in SQL	7
4.1 Descriptive statistics	7
4.2 Time-lagged data engineering.....	7
4.3 Unexpected perfect correlation	7
5. Methodology	8
5.1 Model 1: Household savings VS Future GDP	8
5.2. Model 2: Education VS Savings	8
6. Dashboard Development.....	9
6.1 Geographic Overview Map.....	9
6.2 Regression Explorer.....	9
6.3 5-year Scenario Tool	9
7. Key findings.....	10
8. Policy Recommendations.....	10
9. Limitations	11

1. Executive summary

- European policymakers seek evidence-based strategies to support sustainable long-term economic growth. This project evaluates whether increasing tertiary education attainment or household savings rates is more strongly associated with higher future GDP per capita across European countries.
- Using harmonized data from Eurostat (2013–2023), this analysis integrates household savings rates, tertiary education attainment (levels 5–8), and real GDP per capita (chain-linked volumes, 2020 euros). Data were cleaned and merged in R, validated and explored in SQL, and presented through an interactive Tableau dashboard.
- Key findings indicate that higher household savings rates are strongly associated with higher GDP per capita five years later, explaining approximately 28% of cross-country variation. Tertiary education attainment shows a statistically significant but weaker relationship with savings behavior, explaining about 9% of variation.
- Policy recommendations include promoting household savings through financial incentives and literacy programs, continuing investment in higher education for long-term workforce productivity, and adopting country-specific strategies to reflect diverse economic conditions.
- This analysis identifies statistical associations rather than causal relationships and should be interpreted as an exploratory policy support tool.

2. Business problem

- European countries face varying economic growth trajectories, with some regions experiencing stagnation while others demonstrate sustained expansion. Policymakers require evidence-based insights to determine which socioeconomic factors most strongly support long-term GDP growth.
- While investment in higher education is widely viewed as a driver of productivity, the role of household savings in enabling investment, resilience, and capital formation is less clearly quantified across Europe.
- This project addresses two primary questions:
 1. Is higher tertiary education attainment associated with higher household savings rates?
 2. Are higher household savings rates associated with higher GDP per capita five years later?
- By answering these questions, the analysis aims to inform policy strategies that promote sustainable economic growth.

3. Data Sources and Preparation

3.1 Data sources

All datasets were obtained from Eurostat, the statistical office of the European Union, which provides harmonized and comparable socioeconomic data across member and partner countries.

- The following datasets were used:

Household Savings Rate

- **Dataset:** Household saving rate
- **Eurostat Code:** tec00131
- **Description:** Gross household saving as a percentage of disposable income
- **Coverage:** European countries, 2013–2023
- **Source:** Eurostat

Tertiary Education Attainment

- **Dataset:** Population with tertiary education (levels 5–8)
- **Eurostat Code:** tps00066
- **Description:** Percentage of population with tertiary education attainment
- **Coverage:** European countries, 2013–2023
- **Source:** Eurostat
- **Real GDP per Capita**

Dataset: GDP per capita in chain-linked volumes (2020 euros)

- **Eurostat Code:** sdg_08_10 (CLV20_EUR_HAB)
- **Description:** Real GDP per capita adjusted for inflation
- **Coverage:** European countries, 2013–2023
- **Source:** Eurostat

- All datasets were downloaded as custom extracts to ensure consistent country coverage, time periods, and variable definitions.

3.2 Data cleaning in R

Data preparation was conducted in R using the **tidyverse** package.

Key preprocessing steps:

1. Imported CSV files with missing values encoded as “.” converted to NA.
2. Selected relevant variables (geo, TIME_PERIOD, OBS_VALUE).
3. Renamed variables for clarity:
 - country
 - year
 - edu_rate
 - savings_rate
 - gdp_per_capita
4. Standardized the year variable as an integer to ensure join compatibility.
5. Performed inner joins across datasets by country and year.
6. Removed rows containing missing values.

The use of `inner_join()` ensured that only complete country-year observations were retained.

The final harmonized dataset (`master_data`) was exported for SQL-based validation and exploratory analysis.

4. Analysis in SQL

4.1 Descriptive statistics

- SQL analysis was performed in Google BigQuery to validate and explore the harmonized dataset.

4.2 Time-lagged data engineering

- `LEAD(gdp_per_capita, 5) OVER (PARTITION BY country ORDER BY year ASC)`

- This enabled comparison of current savings rates with GDP outcomes five years later.

4.3 Unexpected perfect correlation

Initial SQL analysis returned a correlation of 1.0 between savings rates and 5-year future GDP in certain subsets. Because perfect correlations are implausible in macroeconomic data, this result prompted further investigation.

The anomaly was attributed to structural artifacts, including:

- Limited 10-year time horizon
- Small sample sizes within lagged subsets
- Potential overfitting within grouped partitions

To ensure statistical robustness, regression modeling was conducted in R. This debugging process improved confidence in the final results.

5. Methodology

5.1 Model 1: Household savings VS Future GDP

A 5-year forward GDP variable was created in R:

- `mutate(gdp_5yr_future = lead(gdp_per_capita, 5))`
- `gdp_5yr_future ~ savings_rate`

Results:

- Coefficient (β): ≈ 1851
- p-value: < 0.001
- R^2 : ≈ 0.28

Interpretation: A 1-percentage-point increase in household savings is associated with approximately €1,851 higher GDP per capita five years later. This represents a statistically significant and moderately strong association.

5.2. Model 2: Education VS Savings

Regression model:

- `savings_rate ~ edu_rate`

Results:

- Coefficient (β): ≈ 0.23
- p-value: < 0.001
- R^2 : ≈ 0.09

- Interpretation: Higher tertiary education attainment is associated with modest increases in household savings rates, though the explanatory power is limited.

6. Dashboard Development

- Dashboard was developed in Tableau to communicate the findings

6.1 Geographic Overview Map

- Toggle between GDP, savings, and education
- Year slider (2013–2023)

Data Challenges

- Country code mismatches (e.g., “EL” vs “Greece”)
- Missing visualization for Montenegro

These were resolved by harmonizing geographic naming within Tableau.

6.2 Regression Explorer

- Scatterplot with regression line
- Year-specific filtering
- Option to display all years (panel view)

Note: Selecting all years increases statistical power but introduces repeated country observations.

6.3 5-year Scenario Tool

- A parameterized slider allows simulated increases in savings rates.

Challenge:

- Initial scaling allowed unrealistic 100% increases.

Solution

Parameter boundaries were set:

- Minimum: 0
- Maximum: 0.1
- Step: 0.01

This constrained projections to realistic 0–10% increases.

The tool is labeled as an exploratory estimate based on historical associations.

7. Key findings

1. Household savings rates show a stronger association with future GDP than education alone.
2. Education contributes indirectly by supporting savings behavior.
3. Economic outcomes vary significantly across countries, suggesting tailored policy approaches.

8. Policy Recommendations

1. Encourage savings through:
 - Tax incentives
 - Financial literacy programs
 - Retirement savings schemes
2. Higher education supports:
 - Workforce productivity
 - Financial stability
 - Long-term economic resilience

9. Limitations

- Correlation does not imply causation.
- GDP outcomes are influenced by variables not included in this model.
- Linear regression assumes stable relationships over time.
- Panel data introduces repeated country observations.
- The scenario slider represents exploratory projections, not forecasts.

Transparent communication of these limitations ensures responsible interpretation.