DWV FINAL

Malak Sayed Ahmed, Faris Alaaeldin El Habashi

ESLSCA 2025

## 1. Dataset Overview

The dataset comprises global education indicators across multiple countries and years, including:

* **Economic metrics**: GDP\_per\_capita
* **Education metrics**: Enrollment rates (primary, secondary, tertiary), Literacy Rate, Avg Years of Schooling
* **Technology & Policy metrics**: Internet Access, Government Policy Score
* **Other**: Gender Parity Index, Unemployment Rate
* **Size:** 10,000 rows

## 2. Data Preparation

Tools: **Python (pandas, numpy, scikit-learn)**

Key steps:

* **Handled missing values** using SimpleImputer
* **Outlier detection** via z-score; handled with **winsorization**
* **Created new features**: e.g., GDP\_per\_capita\_log, winsorized columns
* **Categorical encoding**: Regions grouped; year retained as integer

## 3. Data Transformation

* Created derived tables using **SQL (SQLite)** from cleaned CSVs:
  + region\_avg
  + enrollment\_melted
  + gdp\_combined
* Performed:
  + **Joins** to link GDP and region data
  + **Groupings & aggregations**
* Saved and documented all queries in .sql file

## 4. Exploratory Data Analysis

* **Summary statistics** (mean, median, variance) computed
* **Correlations** analyzed and visualized
* Notable patterns:
  + Higher GDP correlates with higher Literacy and Tertiary Enrollment
  + Regionally, Sub-Saharan Africa lags in GDP and literacy
  + Education levels are rising globally, but disparities remain

## 5. Visualizations in Tableau

Built three dashboards using the exported CSVs:

**Dashboard 1: Global Education Overview**

* **Map of Literacy Rate** by country
* **Bar chart** of top countries by GDP
* Filters for region and year

**Dashboard 2: Trends Over Time**

* **Line chart**: Enrollment rate by education level
* **Animated time slider** to show change from 2000–2020

**Dashboard 3: Regional Insights**

* **Heatmap** of education metrics by region
* **Treemap**: Country GDP breakdown
* **Boxplots**: Literacy by region

*Screenshots to be inserted from Tableau*

## 6. Key Findings

* Economic and education performance are strongly linked
* Some countries have high literacy despite low GDP — due to government policy
* Clustering showed 3 main education+GDP groups globally
* Decision-makers can target regions lagging in tertiary enrollment & internet access