Analyzing the Relationship Between GDP Per Capita and Women’s Tertiary Education Using Data Science Methods

Introduction

This project explores the relationship between economic prosperity, measured by GDP per capita, and the rates of women's tertiary education across OECD countries. The aim is to understand how financial conditions influence educational outcomes for women, with the broader objective of supporting evidence-based policy-making. By combining economic and demographic data, the project provides a comprehensive analysis using statistical techniques and machine learning models.

Datasets were sourced from the OECD and included:

\* Education indicators (percentage of women with tertiary education)

\* GDP per capita (USD)

\* Labour force participation (women’s employment)

\* Fertility rates (births per woman)

After integrating and aligning the datasets by country and year, the project proceeded with exploratory data analysis, hypothesis testing, and machine learning modeling.

Exploratory Data Analysis (EDA)

The initial phase of analysis involved:

\* Merging four OECD datasets on a country-level basis

\* Handling missing values and ensuring numeric data types

\* Generating summary statistics and visualizations

Key EDA findings include:

\* A moderate positive correlation (r = 0.489, p = 0.0021) between GDP per capita and women’s tertiary education levels, which is statistically significant

\* Countries with higher female employment rates also tend to have higher education levels (r = 0.512)

\* A weak positive correlation between fertility rates and women’s education (r = 0.046)

Scatter plots, histograms, and correlation matrices were used to visualize these trends and support further testing.

Hypothesis Testing

Statistical methods were used to validate initial observations:

1. \*\*Correlation Analysis\*\*

\* Pearson correlation between GDP per capita and women’s tertiary education: r = 0.489, p = 0.0021

\* Statistically significant relationship (p < 0.05)

2. \*\*Group Comparison\*\*

\* A t-test comparing education rates in high-GDP vs. low-GDP countries:

\* High-GDP (n=19): mean = 48.60

\* Low-GDP (n=18): mean = 40.53

\* t-statistic = 2.209, p-value = 0.0339

\* Result confirms a statistically significant difference in education rates (p < 0.05)

Machine Learning Methods

Multiple supervised machine learning models were implemented to predict women’s tertiary education rates based on GDP, employment, and fertility data.

1. \*\*Linear Regression\*\*:

\* R² Score: 0.161

\* Mean Squared Error: 129.85

2. \*\*Decision Tree Regression\*\*:

\* R² Score: -0.566

\* Mean Squared Error: 242.17

3. \*\*Support Vector Regression (SVR)\*\*:

\* R² Score: -130.867

\* Mean Squared Error: 20396.89

4. \*\*Random Forest Regression\*\*:

\* Best performing model with R² Score: 0.824

\* Mean Squared Error: 33.38

Cross-validation and hyperparameter tuning were applied to enhance model robustness. Feature importance analysis revealed GDP as the most influential predictor, followed by employment rate and fertility.

Additional Scenarios

\* What-if scenarios were simulated to examine education outcomes under varying GDP levels

\* A \$10,000 increase in GDP per capita corresponds to a meaningful increase in women’s tertiary education, as suggested by model coefficients

Conclusion

This project demonstrates a moderate positive relationship between national economic prosperity and women’s access to higher education. While Random Forest models showed strong predictive power, other models like linear regression underperformed. The analysis supports the idea that investing in economic growth can positively impact gender equality in education.

Future extensions of this project could include:

\* Expanding to non-OECD countries for broader insights

\* Incorporating time-series data for trend analysis

\* Evaluating additional socio-cultural variables

\* Developing interactive dashboards for policymakers

Overall, the study shows that data science methods can effectively quantify complex relationships and aid in shaping informed educational and economic policies.

References

\* OECD Education Database

\* OECD Economic Outlook Database

\* OECD Employment Database

\* OECD Family Database