Capstone Task 3 Report

Western Governors University

Student: Praise Sutton
Program: Data Analytics Capstone
Date: 2025 October

Table of Contents

Deliverable 1.6.4: Policy-oriented Executive Summary Outlining Implications and Recommendations

This project examined the social and economic statistics of fertility decline with a focus on identifying patterns that remain consistent across countries and over time. The results provide clear signals for policymakers seeking to understand the conditions that influence fertility rates.

The analysis finds that marriage is strongly and positively associated with fertility, while higher GDP per capita is linked to lower fertility. These patterns suggest that declining birth rates are not primarily explained by poverty, but rather by shifts in household structures, social norms, and relationship dynamics.

Importantly, the findings run contrary to several common beliefs. Claims that fertility decline is driven by life being "too expensive," by "political instability," or by "women entering the workforce" are not supported by this analysis. None of these factors show consistent, significant correlations with fertility rates.

The implications for policy are significant. Direct financial incentives or "baby bonuses" show little evidence of raising fertility in a meaningful or sustained way. Attention should turn to the quality and stability of family life. Gathering more data around divorce rates, domestic abuse, and dating apps can be the next steps to understanding how the fertility rates are declining. Policymakers should support further research into these topics for causal relationships.

Moving forward, policies that prioritize family and household stability are more likely to be effective. This includes programs that strengthen relationships, reduce domestic stress, and create supportive environments for raising children. By focusing on family well-being rather than short-term financial incentives, governments can better align demographic policies with long term investments in the population.

Part A: Project Overview

The research question guiding this project was: Which macro-level factors most strongly and consistently correlate with national fertility rates? This question addresses an organizational need for policymakers, NGOs, and demographic planners to understand what truly drives fertility decline—whether short-run macroeconomic changes or long-run structural factors such as marriage and income.

The project analyzed global panel data on fertility, marriage, GDP, employment, and political indicators from 1960–2023. Scope was limited to national-level aggregates, excluding individual-level data. The analysis focused on correlations rather than causal modeling, with the goal of ranking the relative strength of associations across multiple dimensions.

The solution applied a structured data analytics methodology: data were compiled into a star-schema database; preprocessing steps included winsorization, logging GDP, and filtering for minimum coverage; analytical methods included pooled correlations, within-

country correlations, partial correlations, and lagged correlations; and tools used were Python (Pandas, NumPy, SciPy, and Matplotlib).

Part B: Project Execution

Originally, the plan emphasized deeper analysis within marital and political status in macro rates. In actuality, the coverage for divorce data and political statistics was weaker than expected, so the focus shifted to GDP and marriage, which yielded robust results. Employment and macroeconomic fluctuations showed minimal correlation with fertility and were treated as secondary.

The project followed a CRISP-DM-style framework with iterative loops as data sparsity forced adjustments. Key milestones included sourcing and schema design, cleaning and preprocessing, pooled correlation analyses, within-country and partial correlations, robustness checks, and final reporting.

Part C: Data Selection and Collection

The dataset was compiled from global, publicly accessible sources, including the World Bank, the United Nations, and Our World in Data. While the initial plan anticipated balanced coverage across all indicators, the actual availability of data varied significantly. Fertility (~94% coverage) and GDP (~81%) were well-represented, while marriage (~68%) and employment (~45%) were moderately covered. Political participation data proved to be the most limited, with only ~2–3% coverage.

During collection, one of the main obstacles was inconsistent reporting across countries and time periods. This required adjusting the scope of analysis to emphasize indicators with broad coverage, while acknowledging gaps in areas such as governance and politics. No major unplanned governance issues arose, though the sparsity of political and employment data restricted the ability to draw strong comparative insights in those domains.

C1. Limitations

Advantages:
- Broad coverage of fertility and GDP enabled strong and reliable cross-country comparisons.
- Data were drawn from reputable, standardized sources, ensuring credibility and consistency.
- The inclusion of multiple domains (fertility, income, marriage, employment, and politics) provided a multidimensional perspective.

Limitations:
- Sparse political and governance data limited interpretability and reduced the scope of conclusions in those areas.
- Moderate to weak coverage in marriage and employment introduced gaps that prevented a fully comprehensive analysis.

- Differences in reporting practices across countries meant that comparisons in certain domains required caution.

Part D: Data Extraction and Preparation

The data were extracted through CSV files and consolidated into a relational database structured in a star schema design, with fertility as the central fact table. Related domains such as GDP, marriage, employment, and politics were organized as dimension tables and connected by ISO codes and year dates, ensuring consistency across sources. The dataset was managed in SQLite, which provided a lightweight but reliable platform for relational joins and efficient querying.

Preparation steps were undertaken to enhance data quality and comparability. GDP per capita was log-transformed to reduce skewness and improve interpretability across countries with vastly different income levels. Outliers were controlled by winsorizing the top and bottom 1% of values, minimizing distortion from extreme cases.

Instead of deleting cases with missing values, data were joined on ISO code and year so that all available observations could be retained. This approach maximized coverage for correlation graphs and comparative analyses, while still ensuring transparency by noting sample size differences across domains.

Part E: Data Analysis

E1. Methods

The analysis focused on examining correlations between fertility and multiple socio-economic and political indicators. Both bivariate correlations and partial correlations (controlling for GDP per capita) were calculated to disentangle overlapping effects. Lagged correlations (1–3 years) were also tested to assess the temporal stability of observed relationships.

E2. Tools and Techniques

All analyses were conducted in SQLite for data organization and Python (pandas, NumPy, and SciPy) for statistical computation. Results were visualized with Matplotlib/Seaborn, which allowed for clear presentation of trends and cross-country comparisons. Winsorization was applied to the top and bottom 1% of values to minimize the impact of outliers.

E3. Analytical Process

To ensure robustness, several checks were performed:

- Minimum-years filters (≥10 and ≥15 years) were applied to confirm that correlations were not driven by countries with very limited longitudinal coverage.

- Lagged models tested whether associations persisted when fertility trends were shifted forward by 1–3 years.

- Comparative interpretation emphasized both the magnitude and direction of relationships across domains.

Key findings indicated that GDP per capita (logged) was the strongest negative correlate of fertility ($r \approx -0.78$), while marriage prevalence was the strongest positive correlate ($r \approx +0.43$). Other domains—employment, politics, and macroeconomic rates—showed weak or negligible associations.

Part F: Evaluation of the Results

F1. Evaluation of Output
The outputs were statistically robust across multiple lenses: pooled, within-country, and robustness checks. GDP and marriage signals were consistent even after winsorization and filtering, while politics, employment, and macro rates faded under stricter tests. This demonstrates that the results are not spurious artifacts of particular transformations.

F2. Value of the Results
The findings provide clear value for policymakers and NGOs. They demonstrate that fertility decline is driven by structural development and household formation rather than short-run macroeconomic shocks. This shifts the policy emphasis away from temporary financial incentives toward long-term investments in relational and institutional stability.

F3. Success Criteria
The project set out to identify persistent, strong correlates of fertility across countries and over time. By consistently highlighting income and marriage as durable signals, the project achieved this goal. Though causal inference was not attempted, the success criteria of identifying reliable correlates were fully met.

Part G: Key Takeaways

G1. Conclusions
1. Fertility declines most strongly with higher income levels, especially when measured as log GDP per capita.
2. Marriage prevalence remains a robust driver of fertility even after controlling for income.
3. Politics, employment, and macroeconomic cycles are secondary or negligible once development is accounted for.

G2. Analytical Tools and Communication
The project used correlations, lagged tests, and partial controls for development, combined with clear textual explanations rather than raw visuals. This approach effectively balanced statistical rigor with accessibility, allowing the results to be communicated to both technical and policy-oriented audiences.

G3. Recommendations
1. Governments should prioritize relational stability and long-term household formation over short-term baby bonuses or financial incentives. Stable family structures have

stronger, more durable associations with fertility.

2. Employers and organizations should invest in sustainable work–life integration and long-term well-being, fostering environments where family formation is supported by stability rather than undermined by precarious work.

Part H: Panopto Recording

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=b4d58e28-57ea-4640-b5e1-b36f0017d82b

Part I: Links

https://github.com/hatodove/Capstone-Project