

Fertility Rates Case Study Rubric

DS 4002 – Spring 2025 - Dana Pham
Due: May 09

Individual Assignment

Why am I doing this? The purpose of this case study is to apply the skills and knowledge you've gained in data science to a real-world problem: understanding the complex social and economic drivers behind global fertility rate trends. By analyzing fertility rates in diverse countries, you'll gain insights that are crucial for addressing global demographic challenges, such as aging populations and the impact of declining birth rates. In doing so, you'll enhance your ability to work with real world data, develop meaningful analysis, and communicate your findings in a way that could inform public policy decisions. This project also allows you to practice important data science tasks such as data cleaning, regression modeling, and visualization, all of which are essential for solving complex, data-driven problems.

What am I going to do? The Github repository for this case study can be found at <https://github.com/dphamm17/CS3-DS4002>. In this case study, you will analyze the relationship between fertility rates and five key social and economic factors—GDP per capita, women's education, life expectancy, child mortality, and homicide rates—across 15 countries over the past 35 years. You will clean and preprocess the data, then apply regression analysis to identify which factors most strongly influence fertility rates. Using your model, you'll make predictions about future trends and create visualizations to clearly present your findings. Finally, you will summarize your results and discuss their implications for policymakers addressing global demographic challenges.

Your final deliverables will include:

- GitHub Repository: Containing code, data, and any resources used for the analysis.

Tips for success:

- Clarity: Ensure that your code and written portions are easy to understand. Make your file names and variable names descriptive.
- Documentation: Comment your code thoroughly so others can follow your thought process.
- Policy Relevance: Keep in mind the potential real-world applications of your analysis, particularly how it could inform policy decisions related to fertility and aging populations.
- Efficiency: Familiarize yourself with the data and tools you'll use early on to complete the analysis more efficiently.

How will I know I have Succeeded? You will meet expectations on Fertility Rates Case Study when you follow the criteria in the rubric below.

| Formatting | <ul style="list-style-type: none">● Repository Structure: Create a new GitHub repository titled "CS_Fertility_Analysis_[Your Name]" that contains: |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|

| | |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | <ul style="list-style-type: none"> ○ README.md: A brief overview of your case study, including the objective, methodology, and results. ○ Data Files: Include the dataset(s) used for the analysis (CSV files, etc.). ○ Source Code: Python scripts or Jupyter notebooks documenting your data preprocessing, regression analysis, and visualization steps. ○ References.md: A markdown file listing all references used (e.g., research papers, online articles) in IEEE citation style. |
| README.md | <ul style="list-style-type: none"> ● <u>Purpose</u>: The README file should provide a clear and concise summary of your project and its key components. It serves as an introduction to anyone viewing your repository and will help them understand the project's goals, methodology, and deliverables. ● <u>Contents</u>: <ul style="list-style-type: none"> ○ Project Title: Include the title of your project at the top. ○ Objective: A brief description of the problem you are addressing (e.g., the decline in fertility rates across 15 countries) and why it matters. ○ Countries and Factors Analyzed: List the 15 countries you are analyzing and the five key social and economic factors being used to model fertility rates. ○ Methodology: Provide a high-level overview of your approach, including data preprocessing, regression analysis, and visualizations. ○ How to Run the Code: Provide instructions for running the analysis, including any setup or dependencies (e.g., Python libraries, packages). ○ References: Mention any sources or studies you referenced to guide your analysis. |
| Source Code | <ul style="list-style-type: none"> ● <u>Purpose</u>: The source code should include all necessary steps to clean, analyze, and visualize the data, allowing others to replicate your analysis. It should be written in Python (or other relevant languages) and be well-documented. ● <u>Contents</u>: <ul style="list-style-type: none"> ○ Data Cleaning: Code to handle missing values and ensure the data is in a usable format for analysis. ○ Exploratory Data Analysis (EDA): Code to calculate summary statistics and visualize the relationships between variables (e.g., scatter plots, histograms). ○ Regression Analysis: Code for fitting and evaluating regression models ○ Prediction and Forecasting: Code to generate predictions for fertility rates based on the regression model. |

| | |
|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | <ul style="list-style-type: none"> o Visualization: Code to create clear visualizations, such as line charts or bar plots, to display key findings. |
| References.md | <ul style="list-style-type: none"> • <u>Purpose</u>: Markdown file titled "REFERENCES.md" The references section should include all external sources that helped inform your analysis and methodology. It should follow IEEE citation style. • Contents: <ul style="list-style-type: none"> o List of References: Include all citations for books, articles, websites, and other resources that contributed to your understanding or approach for the case study. o Format: Follow IEEE citation style for all references. Ensure each citation is complete and correctly formatted. |

Acknowledgements: Special thanks to Professor Rosero for this rubric template!