**Course: CS 478 – Independent Study  
Student: Abdeldjalil Hamidi  
Faculty Supervisor: Professor Daniel Haehn  
Semester: Spring 2025  
Completion Date: May 15, 2025**

### **Introduction**

This independent study focuses on **Investigating Population Collapse**, exploring the factors contributing to demographic decline and its potential long-term consequences. With birth rates declining, aging populations increasing, and migration patterns shifting, many countries face uncertain demographic futures.

Through data-driven analysis, this study will identify key trends, economic and social impacts, and potential policy interventions. The research will rely on historical data, statistical modeling, and case studies to provide a comprehensive understanding of population collapse risks worldwide.

The study is structured into four main phases:

1. **Research & Data Acquisition:** A literature review and data collection from World bank Open.
2. **Data Exploration & Analysis:** Using Python , SQL.
3. **Visualization :** Using Tableau, Pandas.
4. **Final Report & Recommendations:** Summarizing findings, presenting forecasts, and suggesting policy measures.

The research will incorporate machine learning techniques, such as time-series forecasting and predictive modeling, to assess the likelihood of population collapse over the next century. By understanding these trends, policymakers and researchers can develop strategies to mitigate potential societal and economic disruptions.

### **Dataset Descriptions & Sources**

This study utilizes multiple datasets from the World Bank Open Data platform, each containing historical records from 1960 to the present. The datasets are structured with the following headers:

* **Country Name**
* **Years (1960 - Present)**
* **Feature Values (corresponding to each dataset)**

#### **1. Birth Rate Below Replacement Level**

* **Description:** This dataset includes fertility rates for various countries, focusing on whether they fall below the replacement level (approximately 2.1 children per woman).
* **Headers:** Country Name, Years (1960 - Present), Average Fertility
* **Source:** [World Bank Open Data - Fertility Rate](https://data.worldbank.org/indicator/SP.DYN.TFRT.IN)

#### **2. Net Migration**

* **Description:** Tracks the net number of migrants (immigrants minus emigrants) for each country, providing insight into migration trends.
* **Headers:** Country Name, Years (1960 - Present), Net Migration Value
* **Source:** [World Bank Open Data - Net Migration](https://data.worldbank.org/indicator/SM.POP.NETM)

#### **3. Population Age Over 65%**

* **Description:** Displays the percentage of the population aged 65 and older, an important indicator of aging societies.
* **Headers:** Country Name, Years (1960 - Present), Population Age Over 65%
* **Source:** [World Bank Open Data - Population Ages 65+](https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS)

#### **4. Death Rate (Crude)**

* **Description:** Measures the number of deaths per 1,000 people in a given country, useful for understanding mortality trends.
* **Headers:** Country Name, Years (1960 - Present), Death Rate (Crude)
* **Source:** [World Bank Open Data - Crude Death Rate](https://data.worldbank.org/indicator/SP.DYN.CDRT.IN)

#### **5. Population Growth (%)**

* **Description:** Represents the annual population growth rate, calculated based on birth rates, death rates, and net migration.
* **Headers:** Country Name, Years (1960 - Present), Population Growth %
* **Source:** [World Bank Open Data - Population Growth](https://data.worldbank.org/indicator/SP.POP.GROW)

#### **6. Fertility Rate**

* **Description:** The average number of children a woman is expected to have during her lifetime, a key factor in demographic trends and population sustainability.
* **Headers:** Country Name, Years (1960 - Present), Fertility Rate
* **Source:** [World Bank Open Data - Fertility Rate](https://data.worldbank.org/indicator/SP.DYN.TFRT.IN)

#### **7. Birth Rate (Crude)**

* **Description:** The number of live births per 1,000 people in a given year, an important measure of population growth and fertility trends.
* **Headers:** Country Name, Years (1960 - Present), Birth Rate (Crude)
* **Source:** [World Bank Open Data - Crude Birth Rate](https://data.worldbank.org/indicator/SP.DYN.CBRT.IN)

#### **8. Total Population**

* **Description:** The total number of people residing in each country, serving as a foundation for demographic analysis and policy planning.
* **Headers:** Country Name, Years (1960 - Present), Total Population
* **Source:** [World Bank Open Data - Total Population](https://data.worldbank.org/indicator/SP.POP.TOTL)

### **Data Cleaning and Merging (Data\_Merge.ipynb)**

In this section, the process of merging datasets and reshaping them is detailed, forming the foundation for subsequent analysis.

#### **Data Loading and Conversion**

The original datasets were provided in .xls format. Each file was loaded using pd.read\_excel() with the xlrd engine, then converted to .xlsx using openpyxl for compatibility with further data manipulation libraries.

#### **Skipping Unnecessary Rows**

World Bank datasets often contain metadata in the first few rows. The skiprows=3 parameter was used to skip these lines and directly access the data.

#### **Dropping Irrelevant Columns**

Columns like "Indicator Name" and "Indicator Code" were dropped using the drop() method to retain only the numeric data, ensuring clean datasets.

#### **Reshaping Data (Wide to Long Format)**

The melt() function was applied to reshape each dataset from wide to long format, transforming yearly columns into a single "Year" column with corresponding values. This makes the data more suitable for time-series analysis.

Example:

fertility\_long = fertility\_df.melt(id\_vars=["Country Name", "Country Code"], var\_name="Year", value\_name="Fertility Rate")

#### **Data Type Conversion**

To ensure consistent data types, the "Year" column was converted to numeric using pd.to\_numeric(), allowing easier filtering and analysis.

#### **Handling Missing Values**

Missing values were addressed by removing rows where key indicators were missing using dropna(), ensuring the dataset remained clean and consistent.

#### **Merging Datasets**

All datasets were merged using pd.merge() on common columns: "Country Name", "Country Code", and "Year". The merge type used was inner, ensuring that only common records across datasets were retained.

Example:

merged\_df = pd.merge(fertility\_long, birth\_rate\_long, on=["Country Name", "Country Code", "Year"], how="inner")

#### **Final Dataset**

The resulting merged dataset was saved as merged\_data\_fixed\_last.xlsx, providing a unified dataset ready for further exploratory data analysis and modeling.