Project on analysis and visualization of fertility in Europe in 2013 - 2023 - Report

Overview

This report describes the results of the project “Fertility Analysis and Visualization in Europe 2013 - 2023”. The project covers data analysis, cleansing and visualization

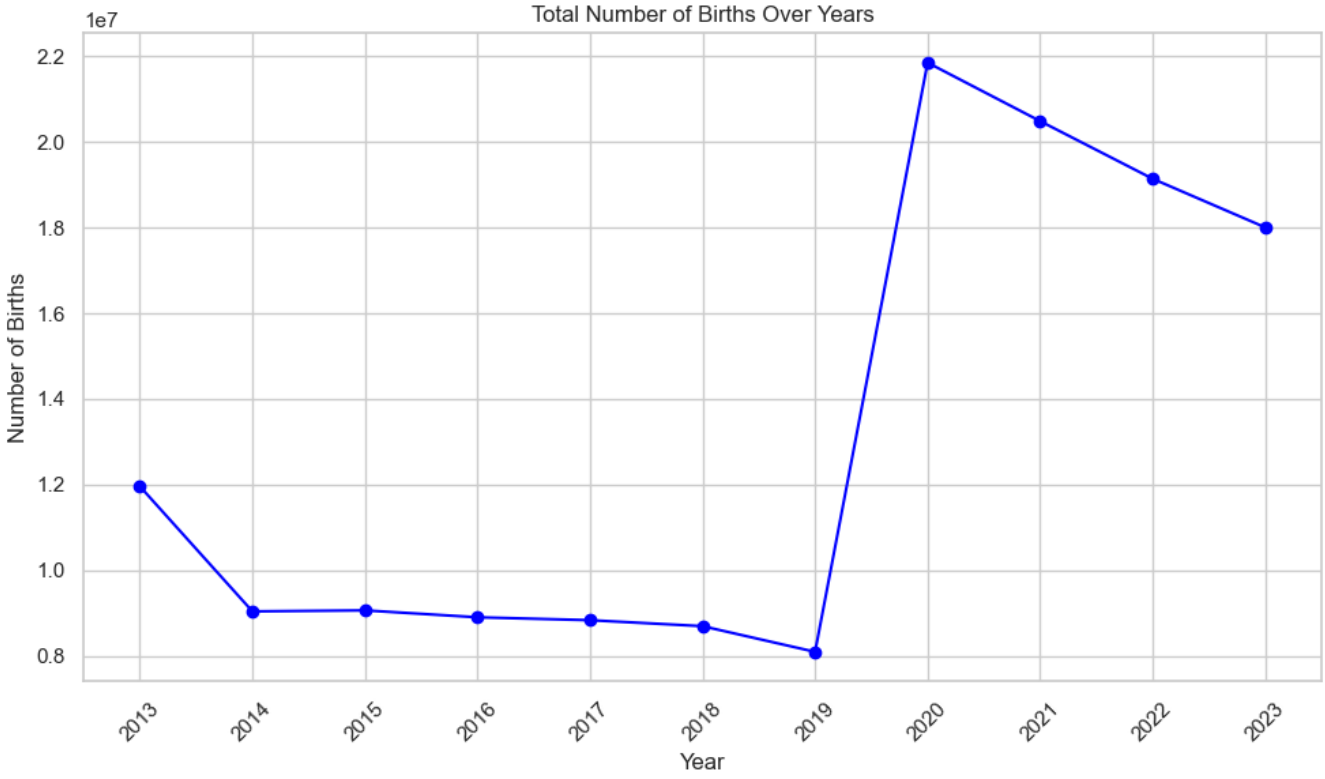
Data cleaning and transformation

- Based on the imperfections of the data obtained for use, the program cleans and transforms them to bring them closer to the “standard normal deviation”
- Outlier removal: A 3-sigma rule to detect and remove outliers that exceed three standard deviations from the mean.
- Log transformation: log1p transformation to reduce skewness and better approximate the normal distribution.

Visual analysis.

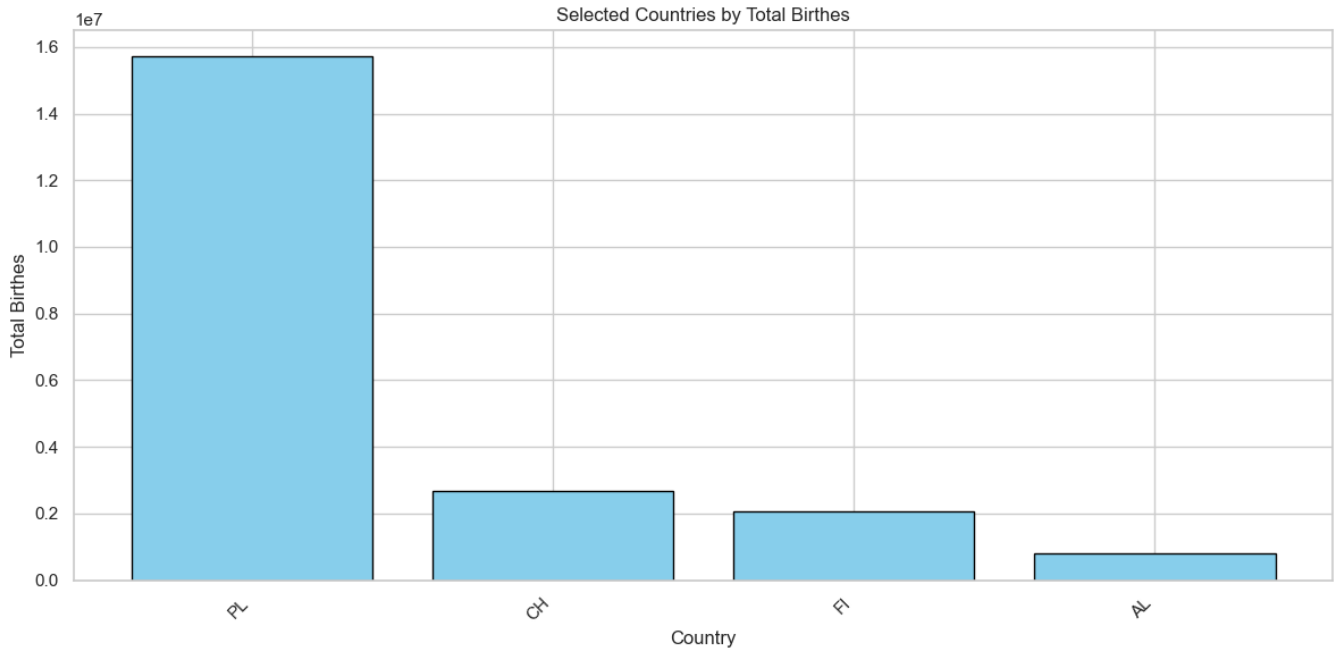
1. Total number of births by year

- The line chart visualizes the general trend of fertility in Europe in 2013 - 2023
- The main observations include periodic increases and decreases that reflect changes in policy, health care or social changes.



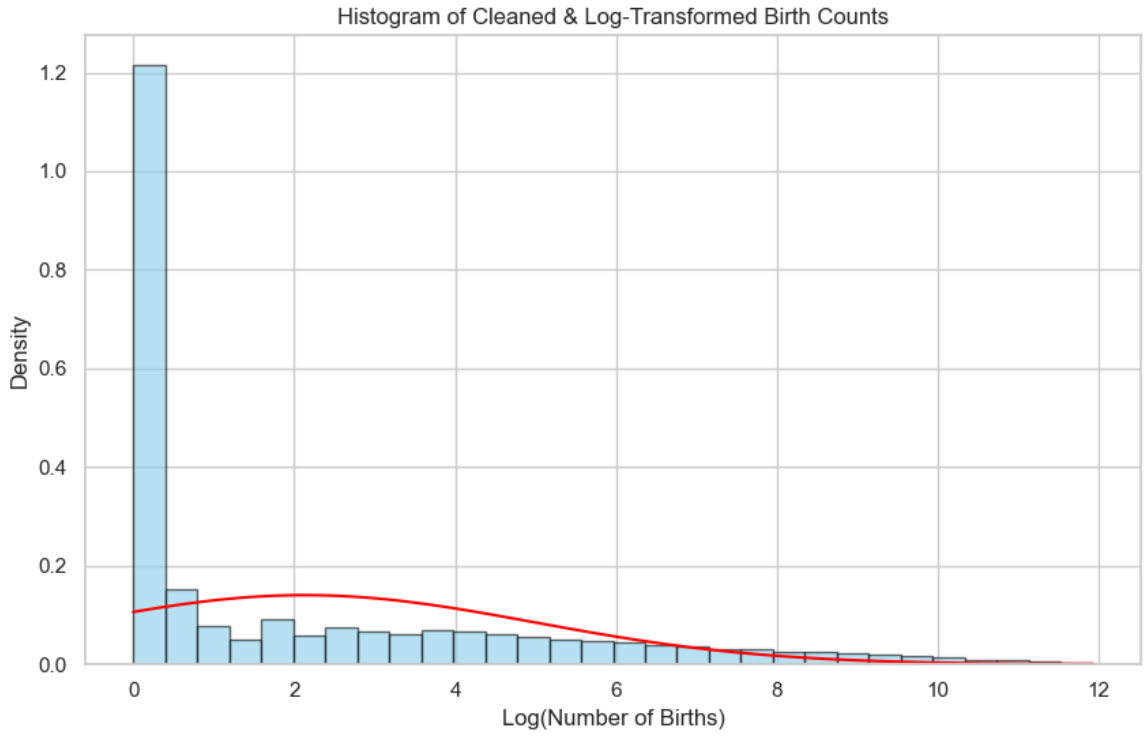
2. Comparison of countries by total number of births

- Comparative visualization allows you to compare the number of births in selected European countries
- It allows you to understand the distribution of fertility at the macro level and identify disparities or unusual trends between countries.



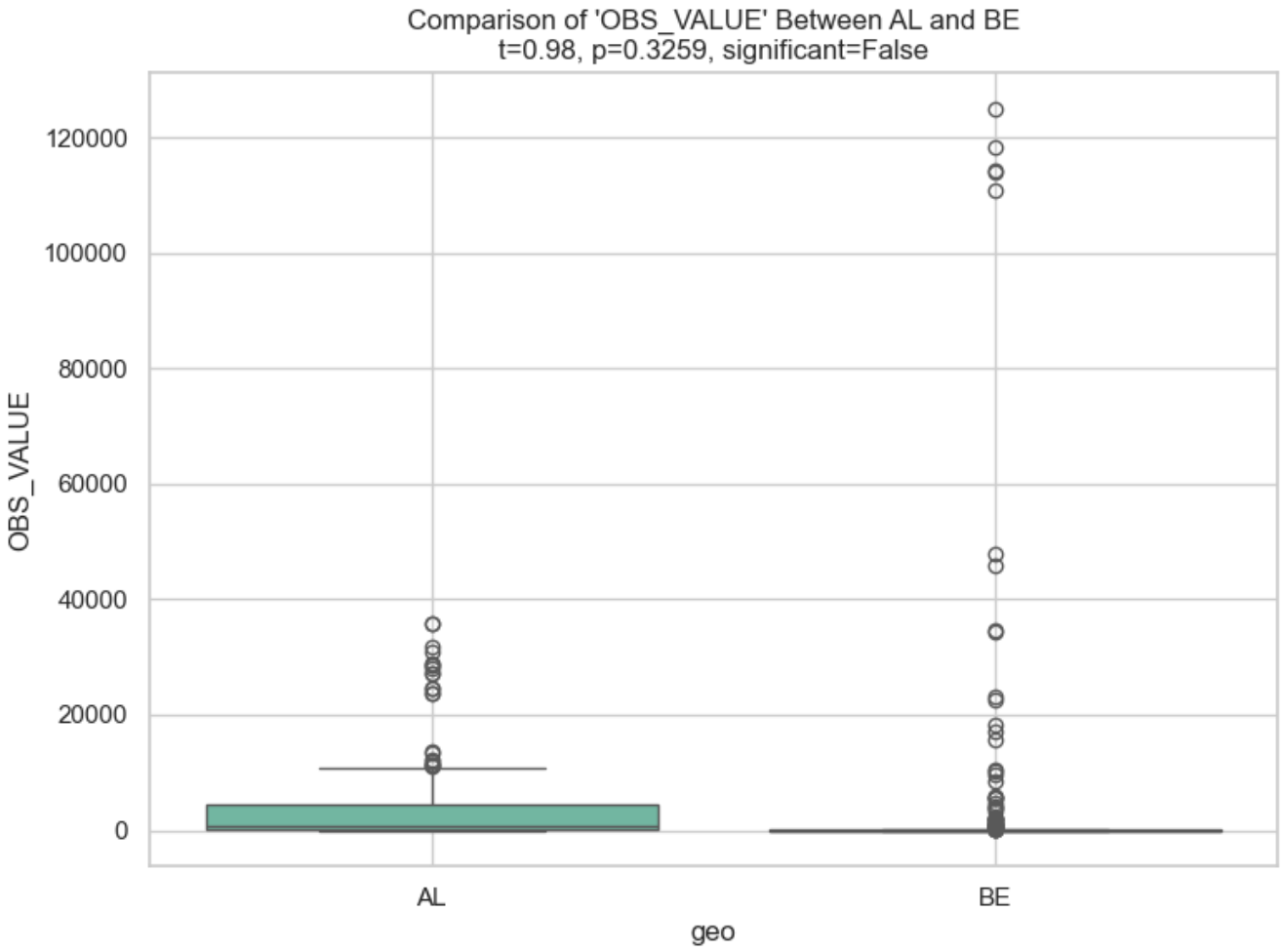
3. Visualization of the overall data distribution

- Visualization of the dataset allows you to clearly understand the overall distribution
- Visualization of the “cleaned dataset” allows you to compare the original picture with the corrected one



4. Hypothesis testing system

- A modular subsystem for hypothesis testing based on user-defined filters (country, direction of comparison, confidence interval).
- Provides p-values and confidence intervals as output.
- Visualization allows you to visually compare hypotheses and helps to better understand their root causes



Statistical Summary

- **Raw Data**: Calculated initial mean and standard deviation for comparison.
- **Cleaned Data**: Recomputed summary statistics post-cleaning.
- **Insights**: Transformation significantly improved data normality, making it more suitable for statistical inference under the **Central Limit Theorem (CLT)**.

Technologies Used

Python 3.9+

- `pandas` — Data wrangling
- `matplotlib` — Plotting
- `numpy` — Math operations
- `scipy.stats` — Distribution fitting
- `InquirerPy` — Interactive CLI