Adrien Protzel
1/15/2025

Birth Rates Across America

Problem Statement: The decrease in birth rates in America can be attributed to a myriad of factors, ranging from economic conditions to cultural shifts. By identifying states with particularly high or low birth rates, we can delve deeper into the unique characteristics and circumstances of these states. This analysis will help us formulate hypotheses about why certain states experience higher or lower birth rates compared to others. Understanding these differences is crucial for developing targeted policies and interventions to address the issue of declining birth rates effectively.

OSEMN Process:

Obtain: The data used in this analysis was sourced from a publicly available dataset provided by Google Cloud BigQuery (“bigquery-public-data.samples.natality”).

Scrub: The data was cleaned by extracting only the states and the number of births. Any rows with missing or incomplete values were excluded from the analysis to maintain data integrity. This cleaning process involved checking for null values.

Explore: The data rows were grouped by states which then compiled the number of births. A bar chart was created to visualize the number of births by state.

Model: The data was processed and visualized using Matplotlib. The states were plotted on the x-axis, and the number of births was plotted on the y-axis.

Interpret: The analysis shows that states with larger populations, such as California and Texas, have higher birth rates. States with lower birth rates may need more resources or a reevaluation of living conditions. This visualization clearly shows which states have the highest birth rates, and similarities and differences can be analyzed to identify key factors contributing to these variations.

Command Line Output:

```
(env) C:\School\Winter-2025-CS512-Data-Science-Tools-Programming\w1>py main-1.py
Running Python 3
Connected to the bigquery client.
Found empty value, skipping it
Generating bar chart
```

Data Visualization: