Future Baby Names Prediction: Leveraging Famous Influences with Machine Learning

An analysis of historical data and trends

YING ZHOU

08/26/2024

Contents

Introduction

A baby's name is a significant gift from parents, embodying their hopes, love, and aspirations for their child.

The process of naming a child is deeply influenced by social experiences, cultural sensitivities, and historical trends.

This project aims to predict popular baby names using machine learning by analyzing historical data from various sources.

Society Value

Cultural Reflection: Names often reflect cultural, social, and generational identities, contributing to a child's sense of belonging.

Historical Significance: Names can capture historical moments, trends, and societal shifts, serving as a record of societal values.

Identity and Aspiration: Parents often choose names that align with their aspirations for their children, influencing future generations.

Business Value

Market Insights: Companies in baby products, toys, and educational services can benefit from understanding naming trends.

Personalization: Brands can personalize products (e.g., custom merchandise) based on popular names in different regions.

Targeted Marketing: Predicting name trends helps businesses tailor their marketing strategies to specific demographic groups.
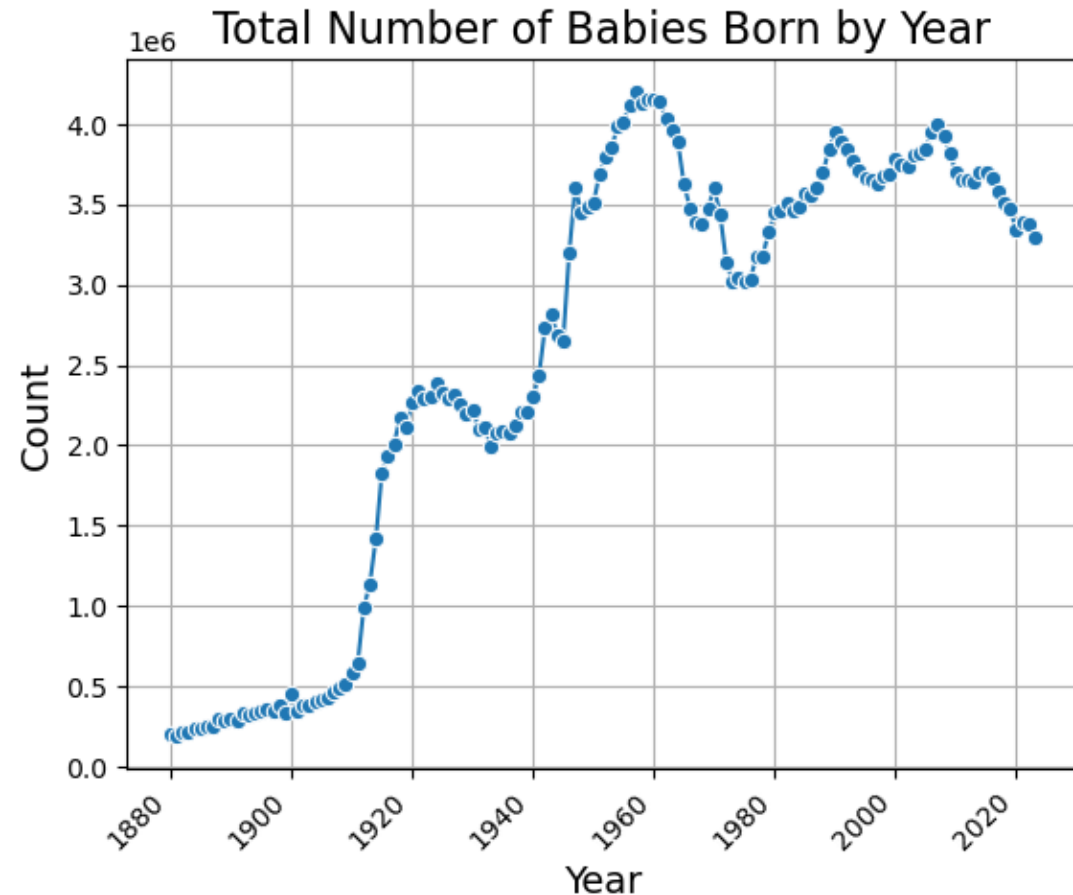
Dataset1: National Baby Name Data (1880-2022)

- Details: *The data was gathered from the United States Social Security Administration's Baby Names dataset. (1880-2022)*

- Dataset variables:

- *Year: Implicit from the filename YYYY.*

- *Name: The name given to babies.*

- *Gender: The gender of the babies ('M' for male, 'F' for female).*

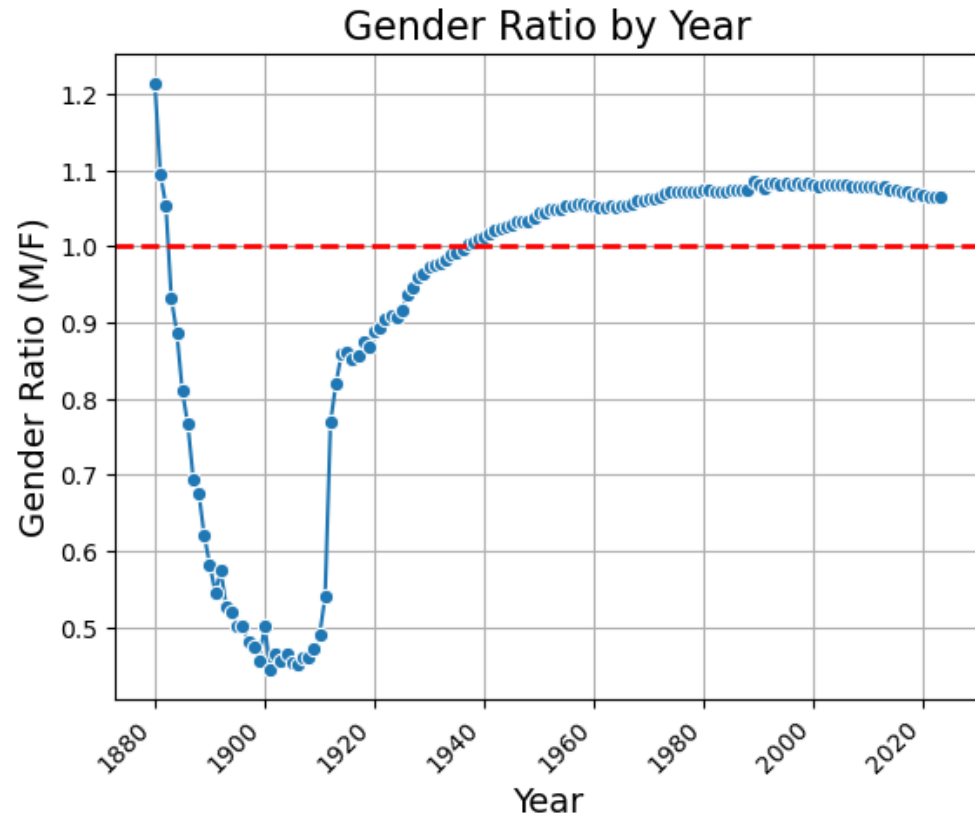- *Count: The number of babies given that name*

Trends in Baby Birth Numbers

Total Number of Babies Born by Year

- **Post-1910 Growth Spike:** Significant societal and economic developments could have contributed to this growth (e.g., industrialization, better healthcare).

- **1960s Decline:** Potential impact of the cultural shifts of the 1960s, introduction of birth control, changing societal roles, and economic factors.

- **Modern Era Stability:** Reflects consistent birth rates due to more effective family planning, economic stability, and societal norms.

Historical Context and Societal Influence on Baby Names and Gender Ratios

Gender Ratio by Year

The Gilded Age (1880s):

Industrial Growth: Exceptional industrial development and economic prosperity.

Railroad Expansion: Increased connectivity across the country.

Political Unrest: Despite economic growth, the era saw violence and political upheaval.

Progressive Era to New Era (1900-1929):

- **Business Expansion:** Growth in businesses and industries.
- **Progressive Reforms:** Social and political reforms aimed at addressing the inequalities and corrupt practices of the time.
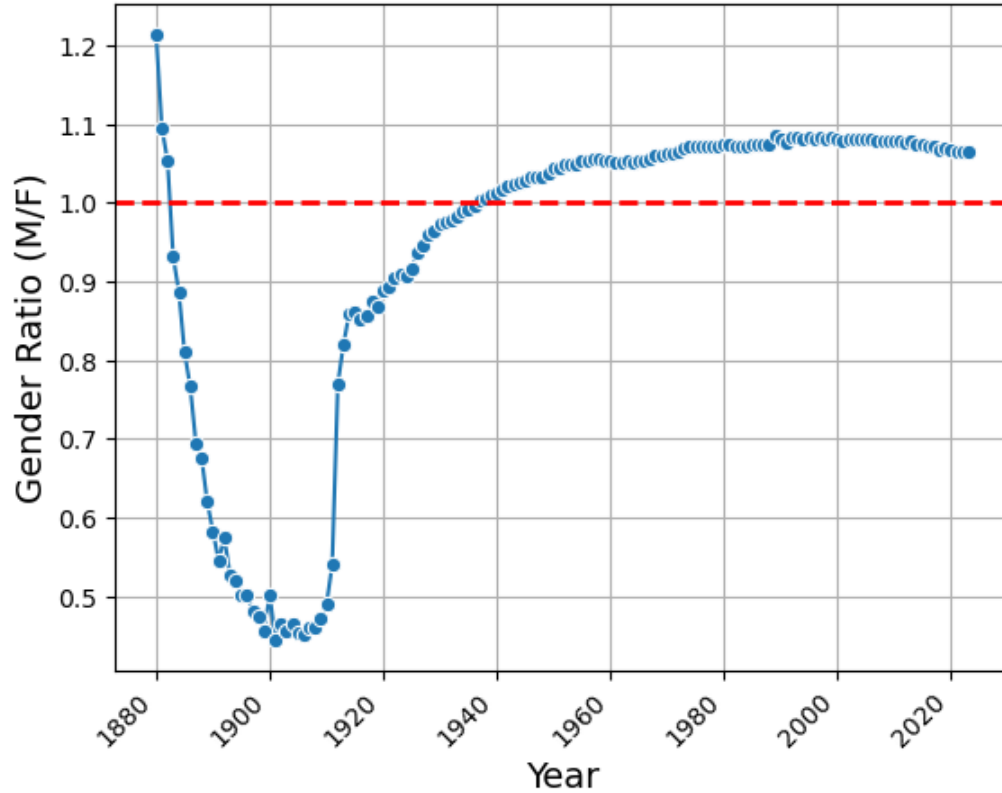
The Great Depression (1929-1939):

- **Economic Downturn:** Severe global economic depression following a sharp decline in stock prices.
- **Impact on Society:** Widespread unemployment, poverty, and social distress.

Historical Context and Societal Influence on Baby Names and Gender Ratios
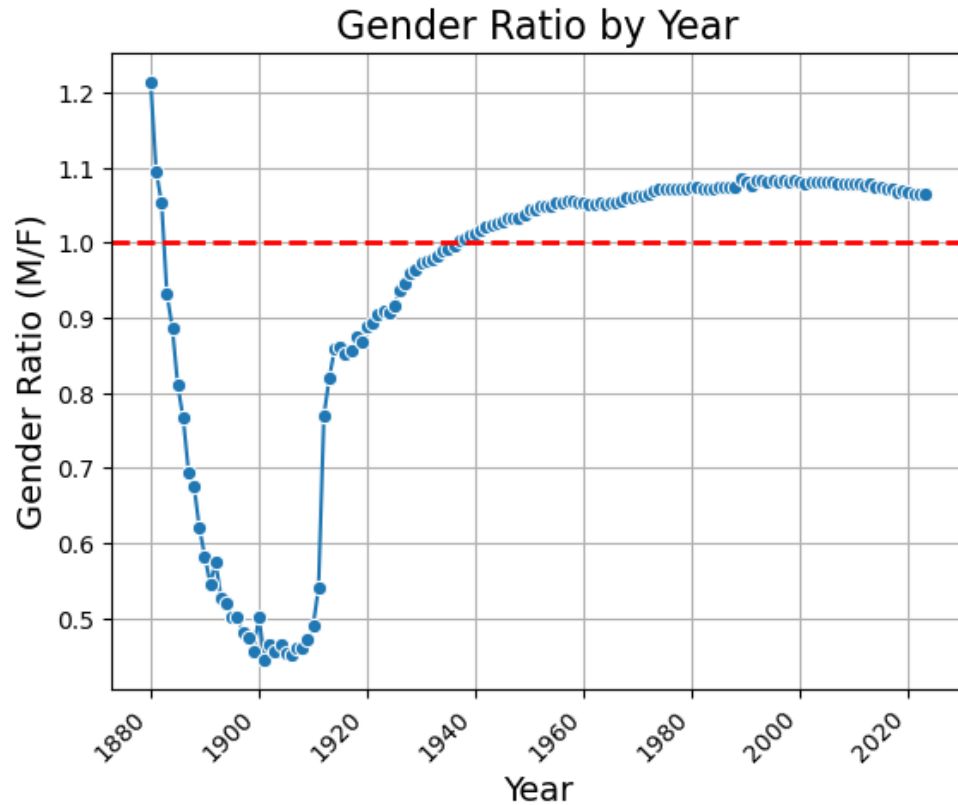

Gender Ratio by Year

Immigration (1880-1930):

- **Large-Scale Migration:** 27-28 million immigrants entered the U.S., primarily from Eastern and Southern Europe.
- **Cultural Diversity:** Significant influence on American culture and naming practices.
- **Mixed Reactions:** Support and opposition, including policies like the Chinese Exclusion Act of 1882.

Urbanization (1870-1920):

- **Rural to Urban Shift:** 11 million people moved to urban areas, with more Americans living in cities than rural areas by 1920.
- **Impact on Naming Trends:** Urbanization brought about cultural integration and diversity in naming choices.

Historical Context and Societal Influence on Baby Names and Gender Ratios

Gender Ratio by Year

Takeaways:

- A trend toward higher female birth ratios during times of economic and societal stress (e.g., Great Depression).

- **Hypothesis:** Evolutionary biology may favor female offspring due to their resilience and better adjustment to adverse conditions.

- **Complexity:** Relationship between societal stressors and gender ratios is influenced by multiple factors, requiring further empirical evidence.
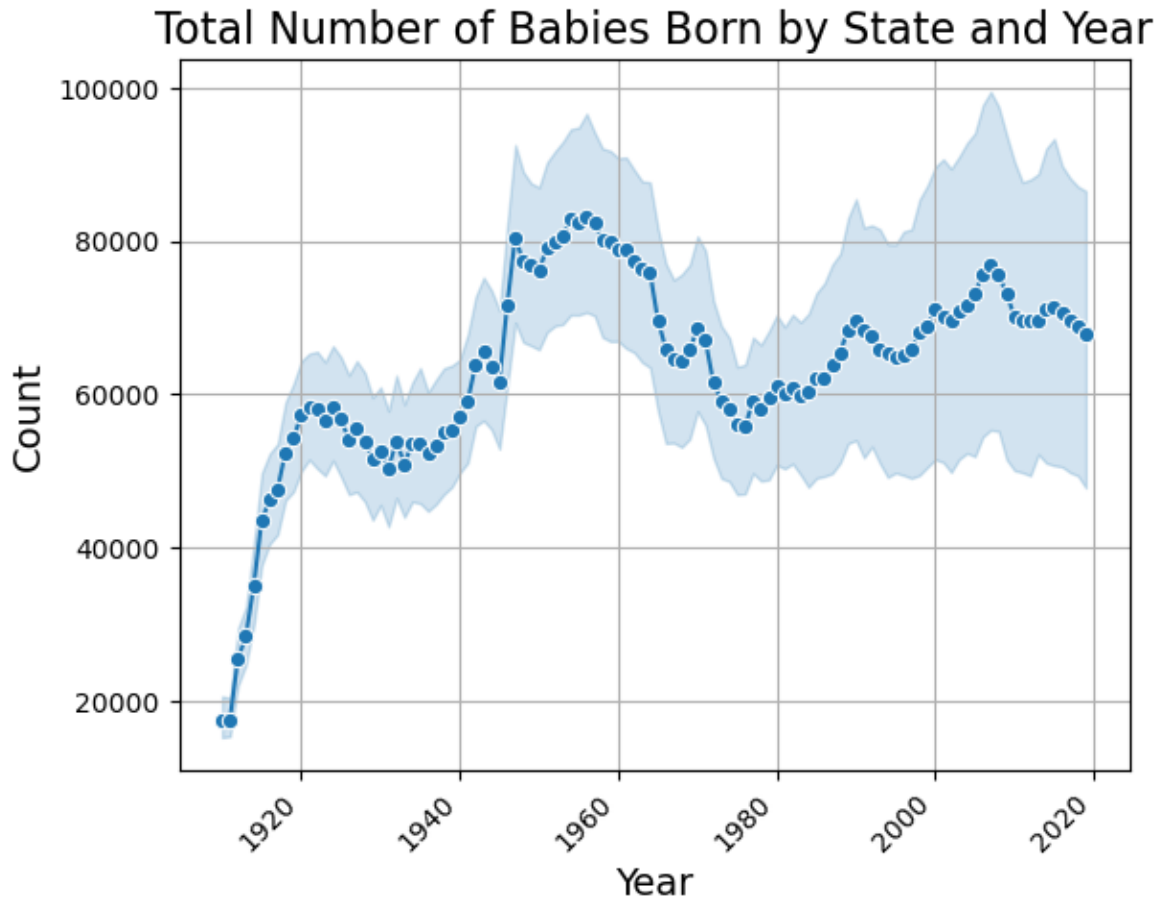
Dataset2: State-Level Baby Name Data (1910-2021)

- Details: Dataset of US baby names from 1910 to 2021. Includes State, Sex, Year, Name, and Count as features.

- Dataset variables:

- *State: The stat.*

- *Year: Implicit from the filename YYYY.*

- *Name: The name given to babies.*

- *Gender: The gender of the babies ('M' for male, 'F' for female).*

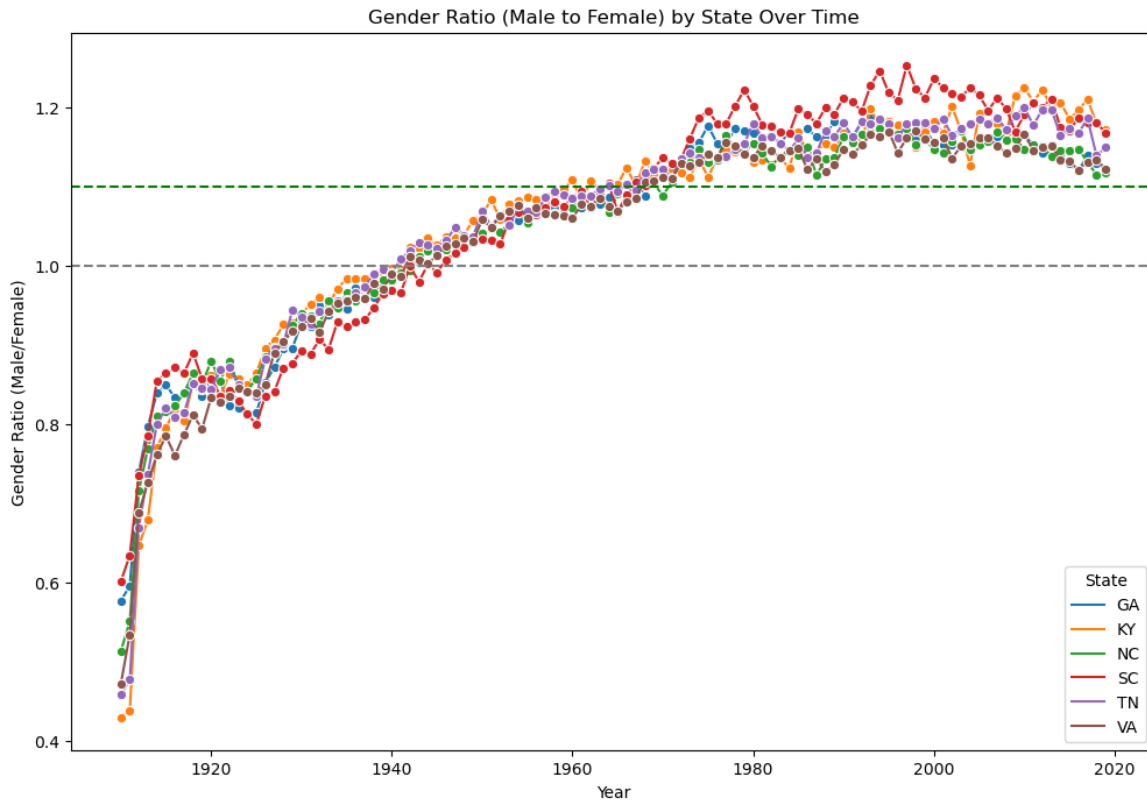- *Count: The number of babies given that name*

Regional Analysis of Baby Birth Trends (1910-2023)

Total Number of Babies Born by State and Year

- **Transition from Homogeneity to Diversity:** While early baby names were influenced less by regional factors, modern trends indicate increased variability and cultural influence on naming conventions.

- **Future Analysis:** Further research into specific cultural and societal events that drive these regional naming differences.

Gender Ratio Analysis in Baby Names (1930-2023)

Gender Ratio (Male to Female) by State Over Time

Takeaways:

- **Increasing Male Ratio:**
    - **Trend:** Since 1930, the male-to-female ratio has been rising.
    - **1970 Milestone:** The ratio reached approximately 1.2, above the national average of 1.1.
    - **Implication:** A notable deviation in male vs. female baby names, indicating a regional trend favoring male names.

- **Regional Focus:**
    - **Location:** Predominantly southern states.
    - **Hypothesis:** Northern regions may show a higher proportion of female births compared to these southern states.
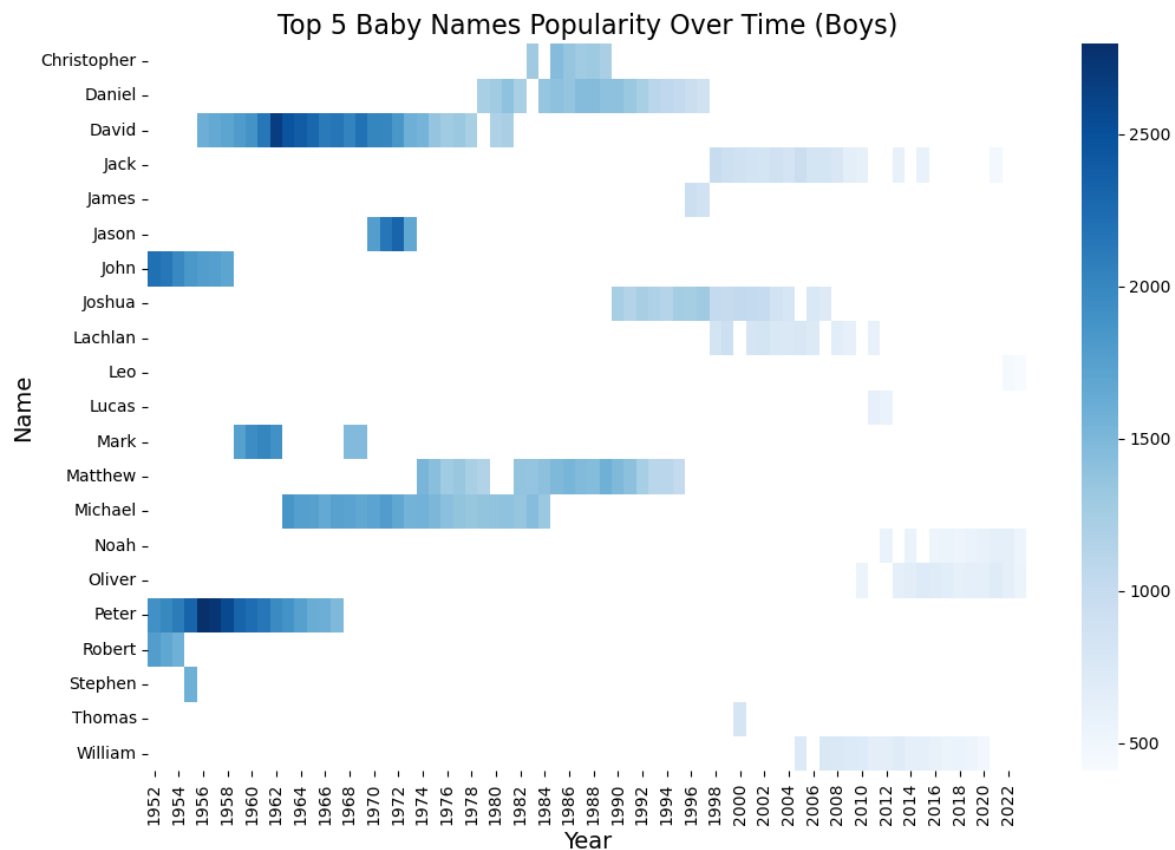
Dataset3: Popular Baby Names (Top 100, 1952-2023)

- Details: *Popular Baby Names from 1952 to 2023 *

- Dataset variables:

- *Rank: The popularity of the baby name in each year.*

- *Year: Implicit from the filename YYYY.*

- *Name: The name given to babies.*

- *Gender: The gender of the babies.*

- *Number: The number of babies given that name*

Trends in Top 5 Baby Boy Names Over Time

Top 5 Baby Names Popularity Over Time (Boys)

- **Long-Term Popularity vs. Brief Spikes:** Some names maintain popularity over many years, while others are only briefly popular.

- **Increase in Newborns:** There has been a significant rise in the number of newborns, especially post-2000.

- **Trend Toward Greater Name Diversity:** Despite the increase in total births, the popularity of individual top names has decreased, indicating a shift towards a wider variety of names.
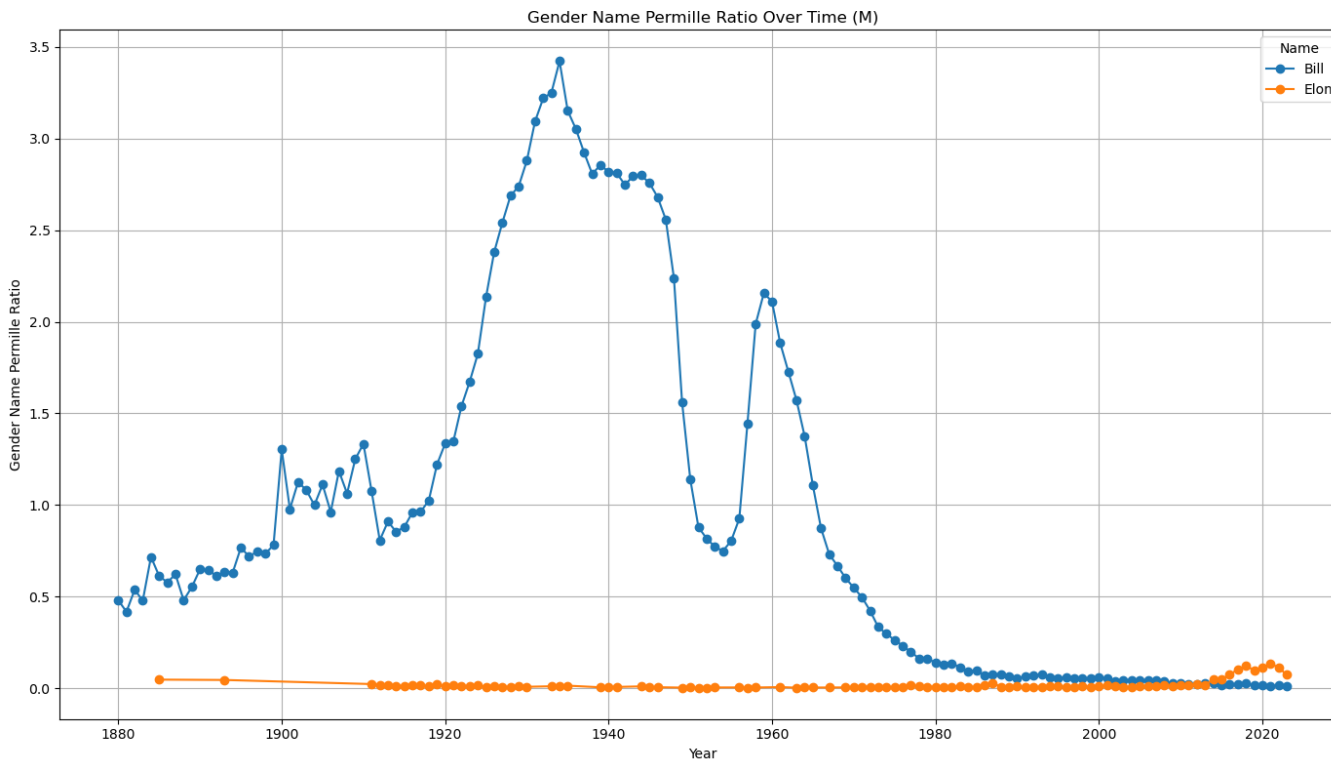
Dataset4: Famous People Names

- Details: *This dataset includes all personal names listed in the Wikipedia category "American people by ethnic or national origin" and all subcategories fitting the pattern "American People of [ ] descent", in total more than 25,000 individuals. Each individual is represented by a row, with columns indicating binary membership (0/1) in each ethnic/national category.*
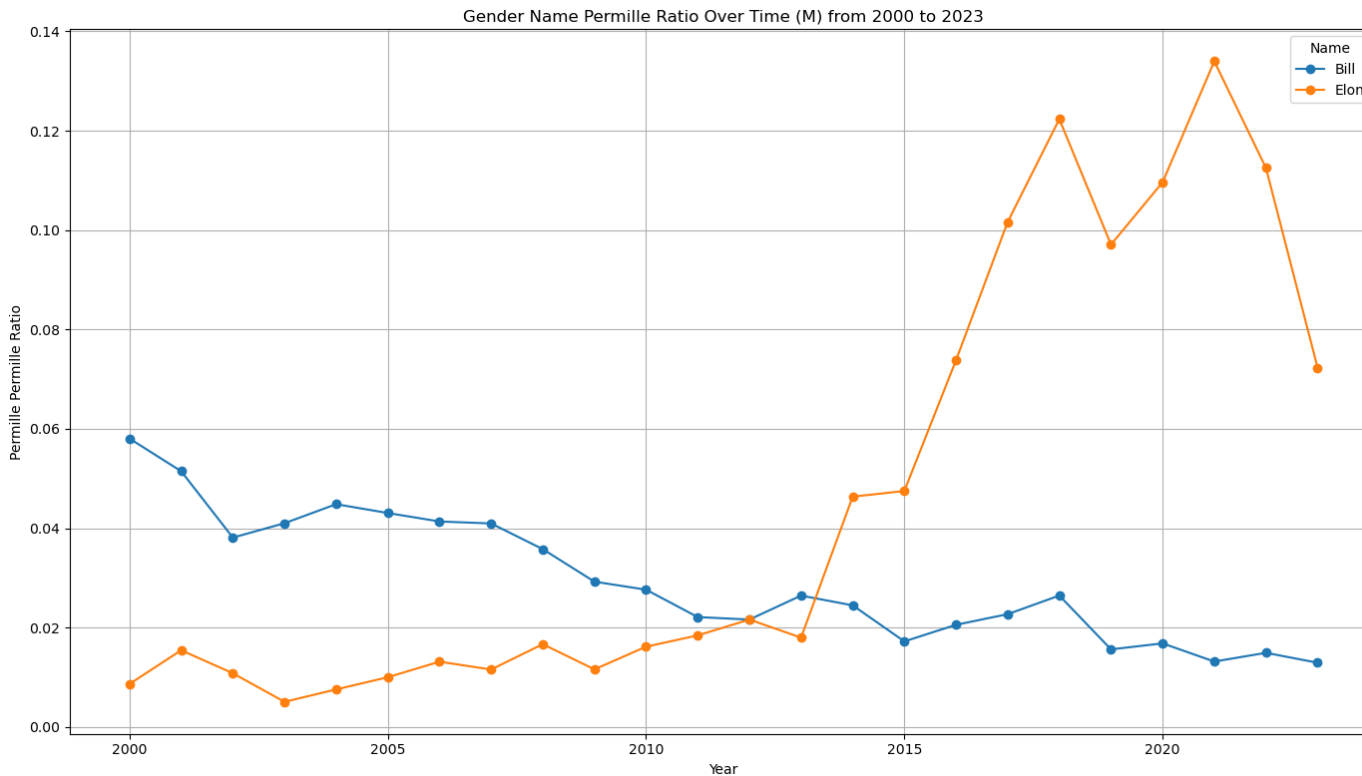
Influence of Famous People on Baby Names

Gender Name Permille Ratio Over Time (M)

Takeaways:

- **Impact of Fame on Name Popularity:** *Elon Musk:* Gained significant fame around May 2012, leading to a more than tenfold increase in the popularity of the name "Elon" compared to previous years.

- The name was rarely used before Musk's rise in fame.

- **Delayed Impact:** The influence of a public figure on name popularity often shows a lag due to the natural delay of pregnancy and birth.

- The spread of reputation and name adoption takes time as the public becomes more familiar with the figure.
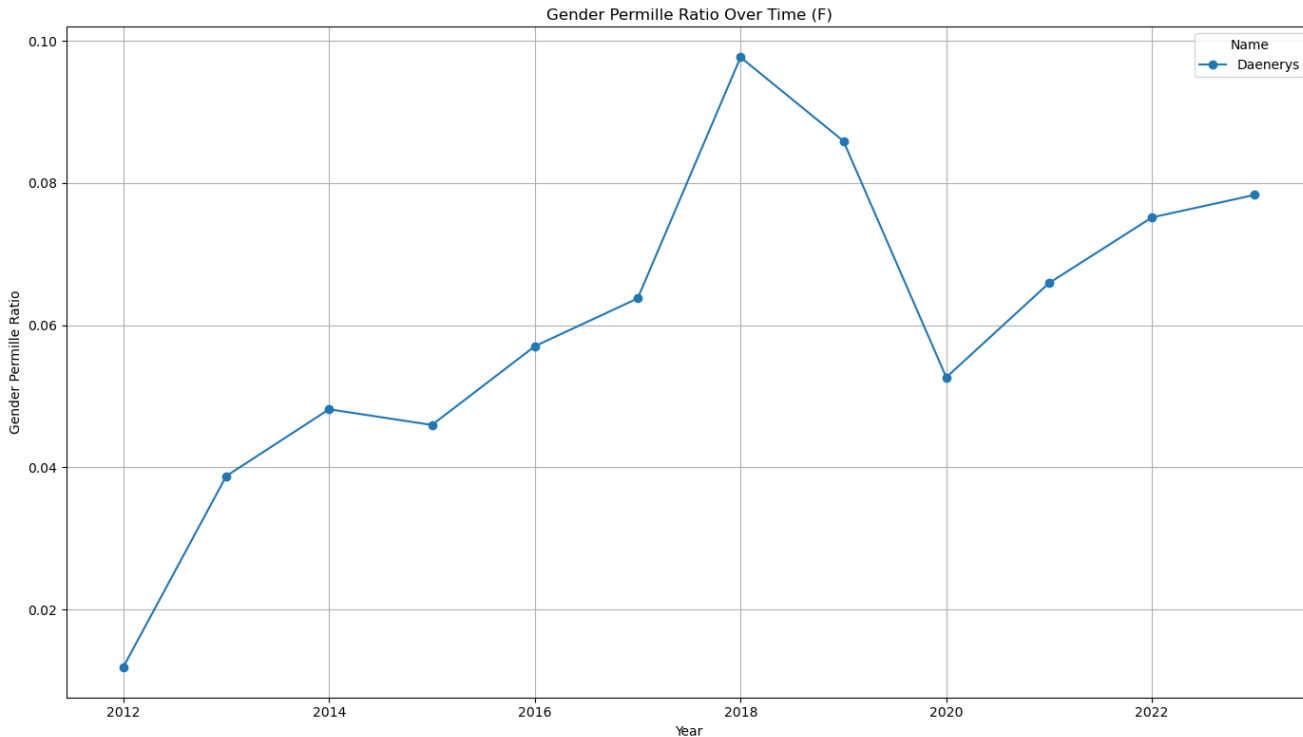
Influence of Famous People on Baby Names

Gender Name Permille Ratio Over Time (M) from 2000 to 2023

Takeaways:

- **Impact of Fame on Name Popularity:** *Elon Musk:* Gained significant fame around May 2012, leading to a more than tenfold increase in the popularity of the name "Elon" compared to previous years.

- The name was rarely used before Musk's rise in fame.

- **Delayed Impact:** The influence of a public figure on name popularity often shows a lag due to the natural delay of pregnancy and birth.

- The spread of reputation and name adoption takes time as the public becomes more familiar with the figure.

Game of Thrones on Baby Names: The Case of "Daenerys"

Gender Permille Ratio Over Time (F)

Initial Rise in Popularity:

• The name "Daenerys" saw a delayed but steady increase as the character gained fame in the early seasons of *Game of Thrones*.

• The popularity of the name grew as the show progressed, reflecting the cultural impact of the character.

Impact of Season 8:

• Following the controversial events in Season 8, where Daenerys's character arc disappointed many fans, the popularity of the name experienced a decline.

• Despite the drop, the name "Daenerys" remained relatively popular, showcasing the lasting cultural significance of the character.

Demo Platform

App Features:

Interactive Exploration: Users can explore historical name trends through an intuitive app interface.

Name Prediction: The app predicts future popular baby names based on historical data and current trends.

Demonstrate the app's functionality, showcasing key features like name search, trend analysis, and prediction outputs.

Next Steps

1

Data Modeling: Using machine learning, like time series to predict the future baby names.

2

Advanced Modeling: Improve predictive accuracy with advanced machine learning techniques and real-time data updates.

3

User Testing: Conduct user testing for the demo platform to refine the user experience and add new features.

THANKS FOR YOUR LISTENING!