

# Baby Names Machine

---

**Future Baby Names  
Prediction: Leveraging  
Famous Influences with  
Machine Learning**



**YING ZHOU**

**09/27/2024**

<https://www.parents.com/top-baby-names-2023-8644280>

<https://www.eventbrite.com/e/childbirth-preparation-class-la-porte-tickets-450644990297>

# Contents

---

---

INTRODUCTION

---

RESEARCH VALUE

---

METHODOLOGY

---

RESULTS

---

DEMO PLATFORM

---

NEXT STEPS

---

CONCLUSION



# 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 historical data and machine learning.



# Research Value

---

---

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

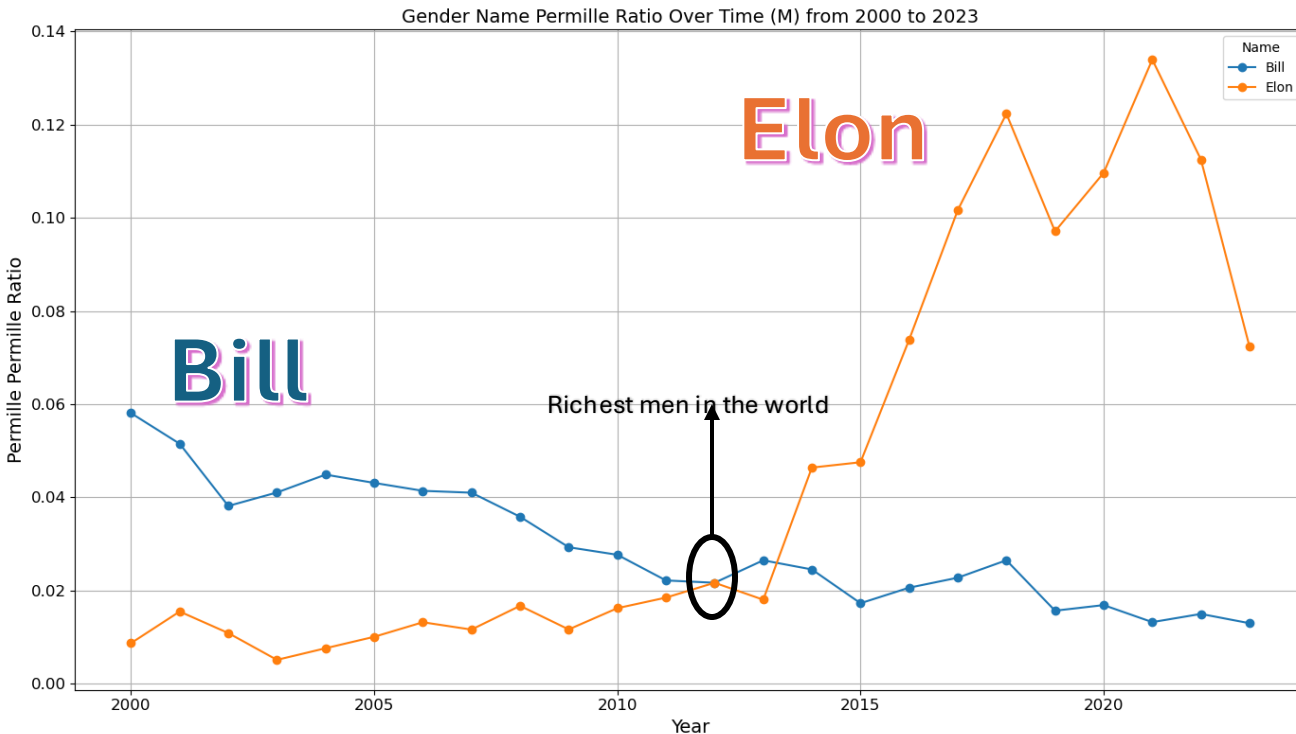
---

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

---

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

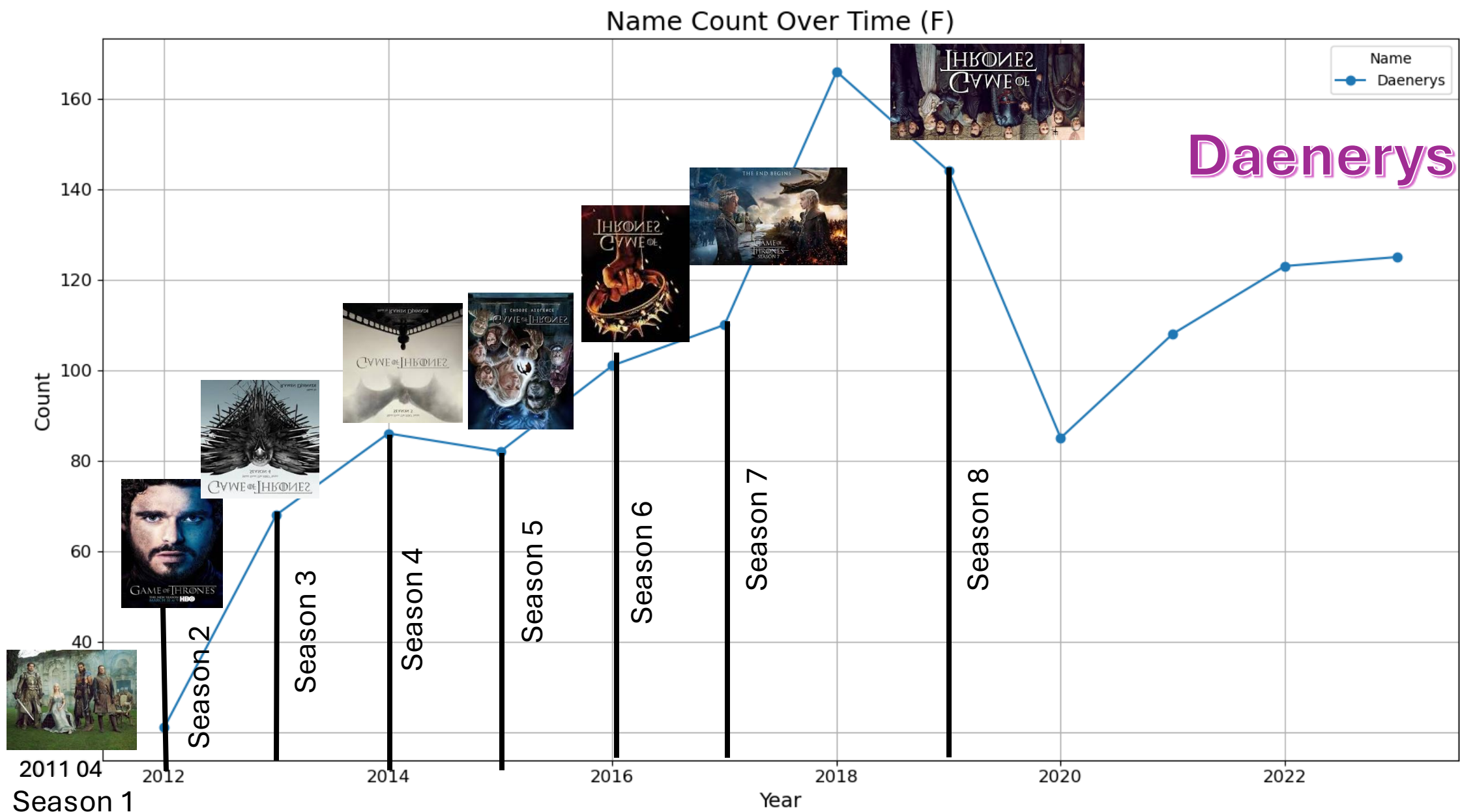
# Methodology



## Takeaways:

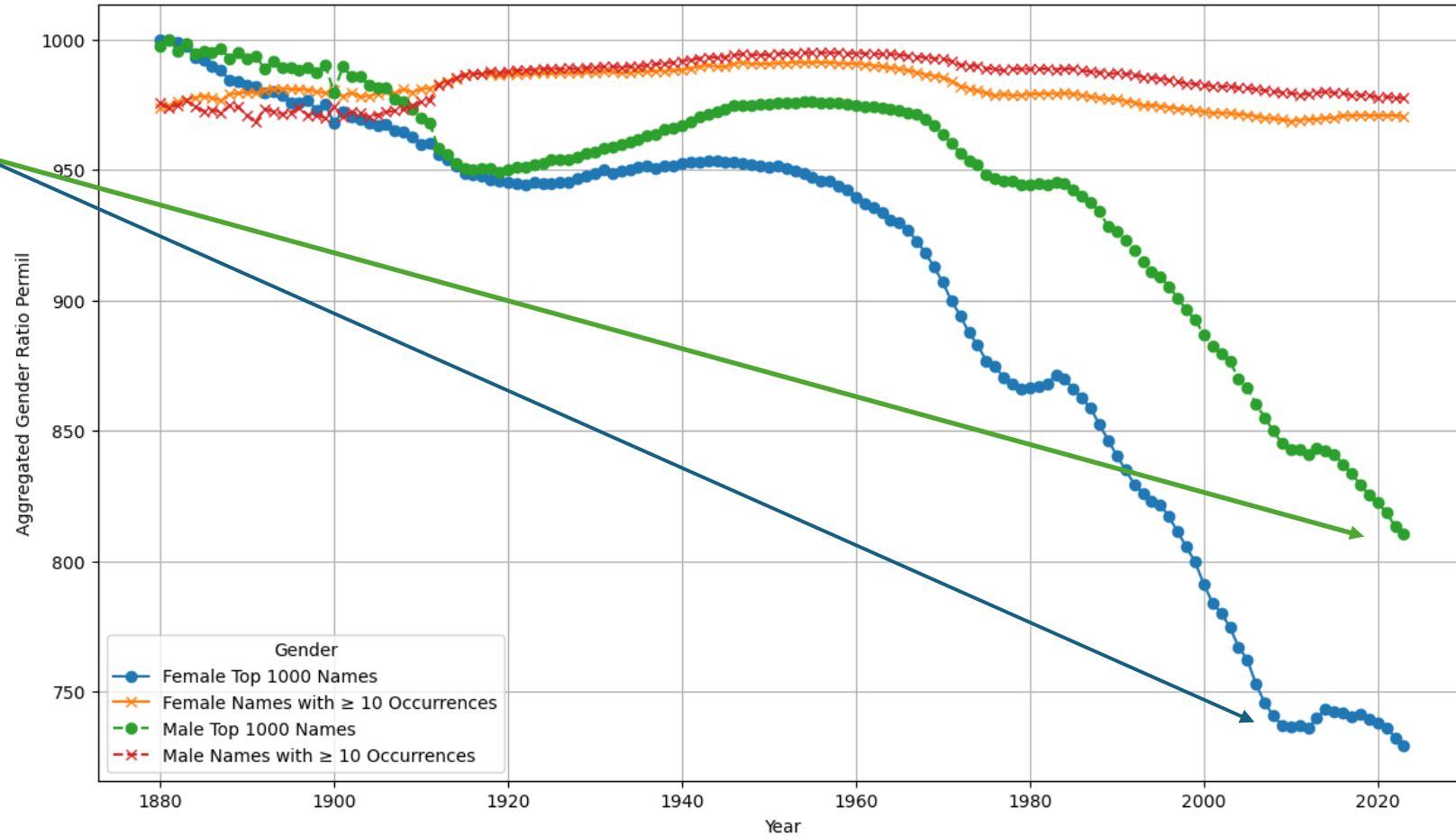
- **Impact of Fame on Name Popularity:** *Elon Musk*: Gained significant fame around May 2012, leading to a more than tenfold increase in the popularity of the name "Elon" compared to previous years.
- The name was rarely used before Musk's rise in fame.
- **Delayed Impact:** The influence of a public 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.

# Methodology



# Top 1000 names vs. Name Occurrences More Than 10 Times

Comparison of Gender Ratio Permil Aggregation Over Time

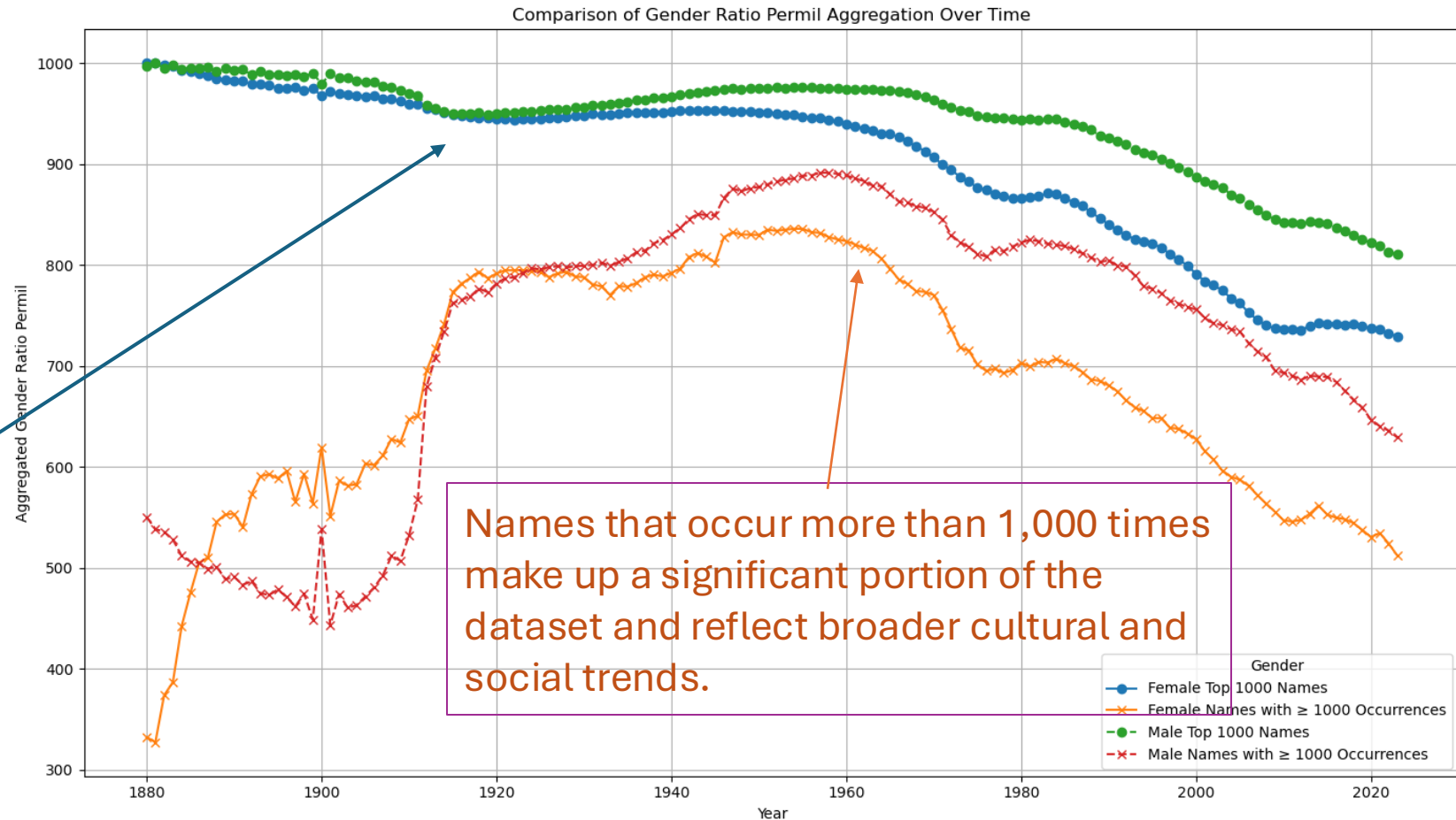


Top 1000 names:  
the majority of the  
population



# Top 1000 names vs. Popular Name Occurrences More than 1000 times

The top 1000 names provide a **reliable sample**, capturing enough **diversity** and **frequency** to ensure the model is both **scalable** and **accurate**.





# Dataset



## Basic Feature

- Gender\_Binary
- Name
- Year

## Name Features:

- Name\_Length
- Vowel\_Count
- Vowel\_Ratio
- Ends\_With\_Specified\_Letters (a, e, i)

## Other Features:

- Year\_of\_Last\_Appearance
- Is\_Famous
- Is\_Top\_100

## Count Features:

- Count
- Name\_Ratio
- Gender\_Name\_Ratio

## Yearly Changes Features:

- Yearly\_Change\_Count
- Yearly\_Change\_National\_Ratio
- Yearly\_Change\_Gender\_Ratio

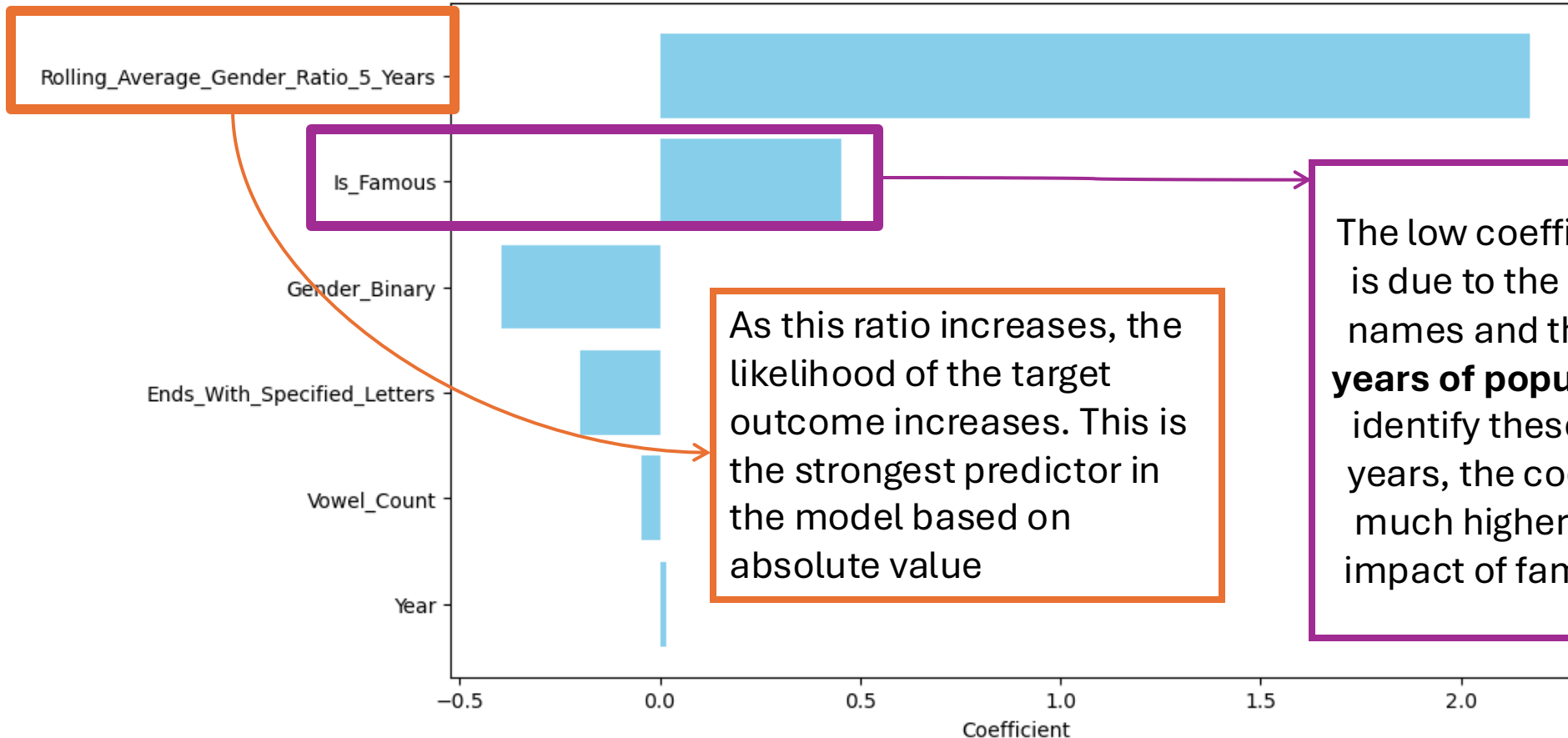
## 5 Year Window Features:

- Rolling\_Average\_Count\_5\_Years
- Rolling\_Average\_Gender\_Ratio\_5\_Years
- Rolling\_Average\_National\_Ratio\_5\_Years

# Logistic Regression Model Coefficients



The Most Important Features



As this ratio increases, the likelihood of the target outcome increases. This is the strongest predictor in the model based on absolute value

The low coefficient for famous names is due to the limited input of famous names and the **lack of data on their years of popularity**. If we could better identify these names and their peak years, the coefficient would likely be much higher, reflecting the stronger impact of fame on baby name trends

# Model Performance

- **Training Accuracy: 0.9787**
- **Test Accuracy: 0.9779**



## Classification Report

### Confusion Matrix

25454	232
405	27818

### Precision

### Recall

### F1-score

0	0.98	0.99	0.99
1	0.92	0.87	0.90

# Demo

## Baby Name Trends

Enter names (comma-separated):

Ayden, Aiden

Enter start year:

2000

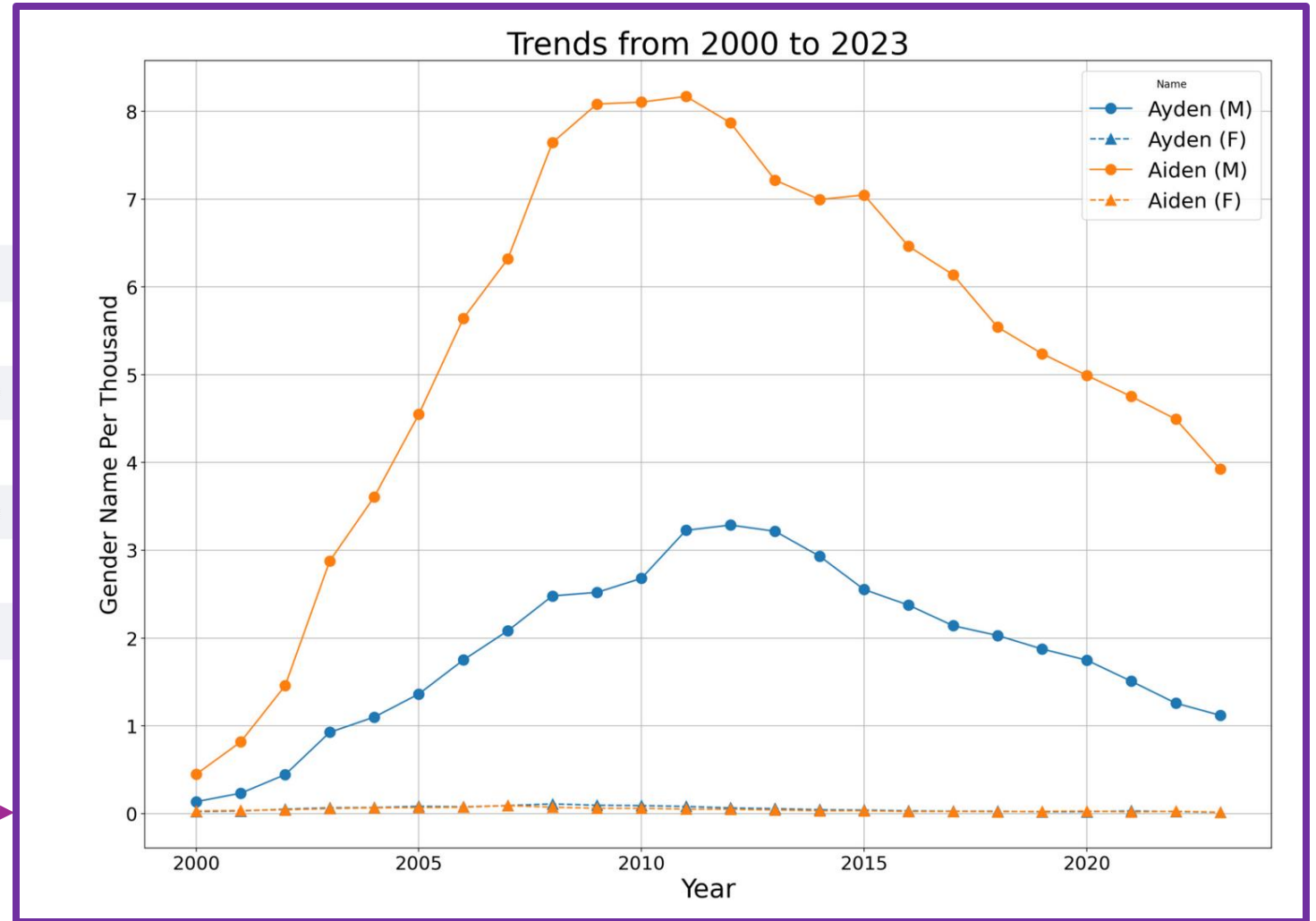
Enter end year:

2023

Select metric to plot:

Gender\_Name\_Ratio

Show Trend



# Demo

## Baby Name Prediction

Ayden

Year

1880

2024

Is the Name Famous? (1 = Yes, 0 = No)

0

Gender (1 = Male, 0 = Female)

1

Rolling Average Gender Ratio (5 Years)

1.50

-

+

Number of Vowels

2

10

Ends with Specified Letters (1 = Yes, 0 = No)

0

### Prediction Results

The name is predicted to be **not** in the Top 100.

Probability of being in the Top 100: 0.000

0

## Baby Name Prediction

Aiden

Year

1880

2024

Is the Name Famous? (1 = Yes, 0 = No)

0

Gender (1 = Male, 0 = Female)

1

Rolling Average Gender Ratio (5 Years)

4.70

-

+

Number of Vowels

3

10

Ends with Specified Letters (1 = Yes, 0 = No)

0

### Prediction Results

The name is predicted to be **in** the Top 100.

Probability of being in the Top 100: 0.904

0.9

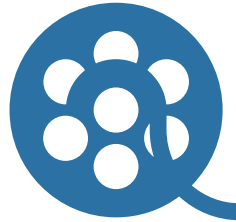
# Next Steps

Social



**Adding Name Meanings**

Tech



**Time Series Analysis for each  
name popular life-time.**

User



**Deep Learning for Name  
Generation**



# Conclusion

- This project showcases how we can use machine learning to predict baby names that balance popularity and uniqueness, helping parents navigate naming decisions. By incorporating trends influenced by culture, famous figures, and historical data, we can better understand how names evolve over time.





# THANKS FOR YOUR LISTENING!

---

- [ying6.zhou@gmail.com](mailto:ying6.zhou@gmail.com)
- <https://github.com/yinger921>

# Supporting Information

# 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.

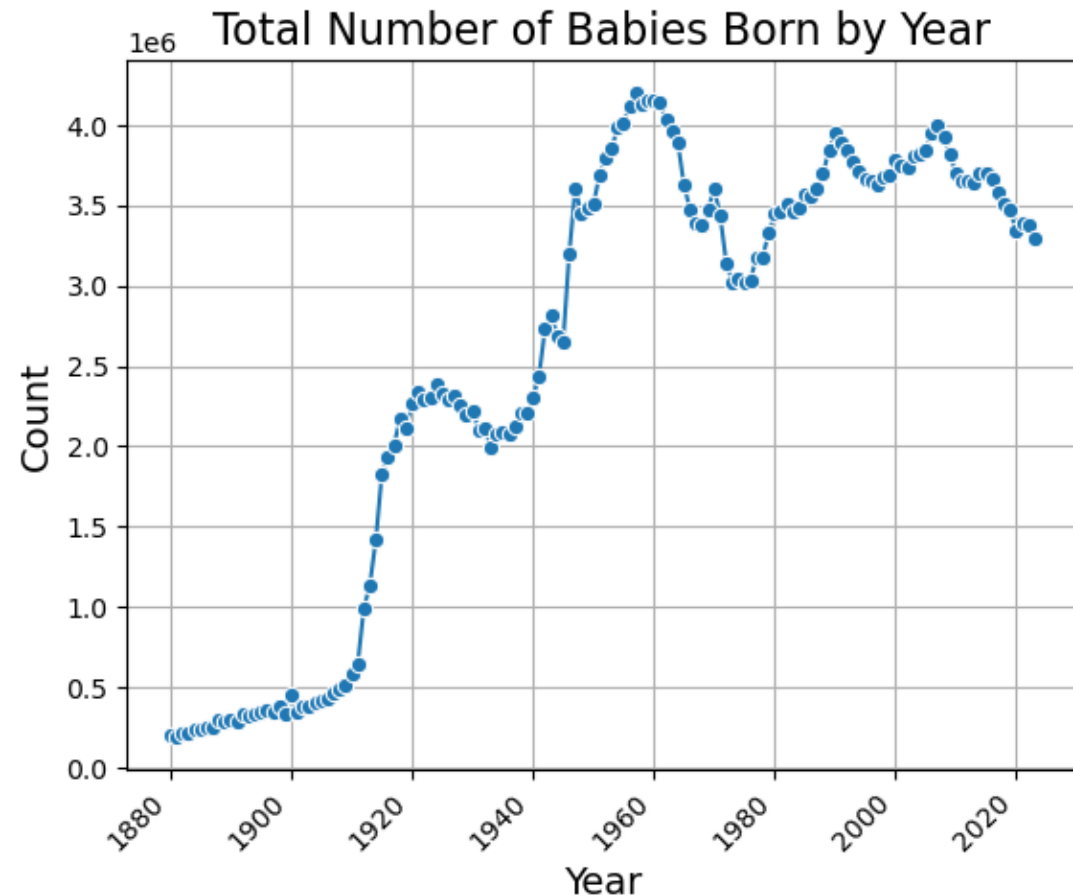
**Targeted Marketing:** Predicting name trends helps businesses tailor their marketing strategies to specific demographic groups.

# 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*

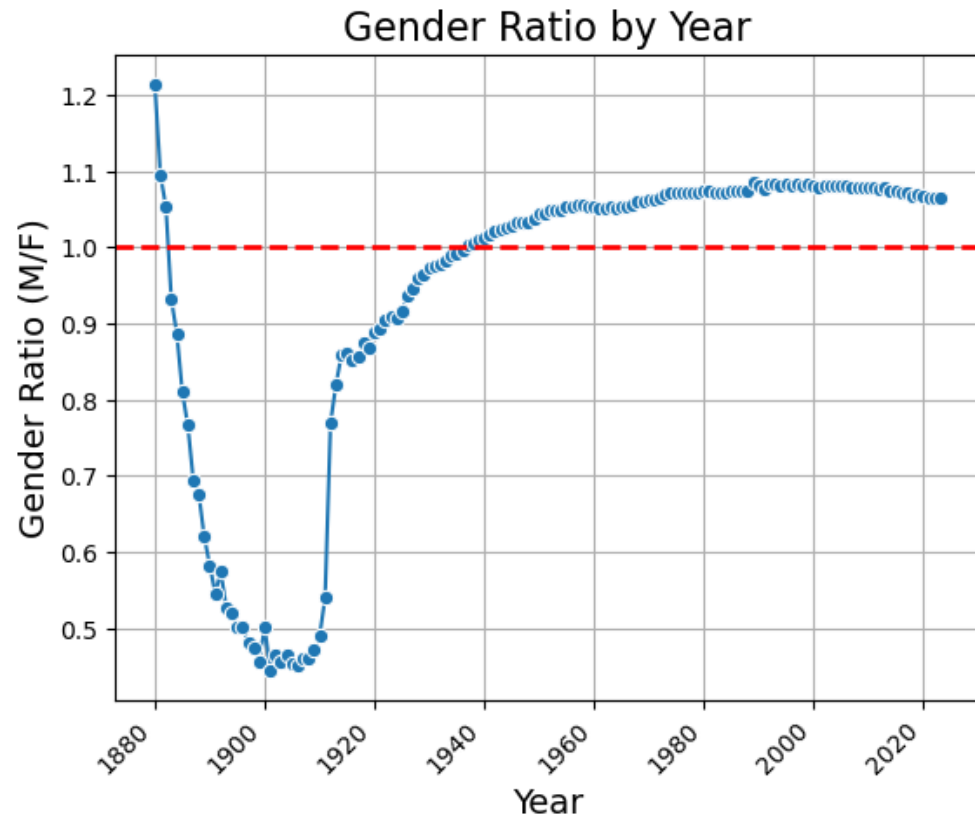


# Trends in Baby Birth Numbers



- **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



## 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):

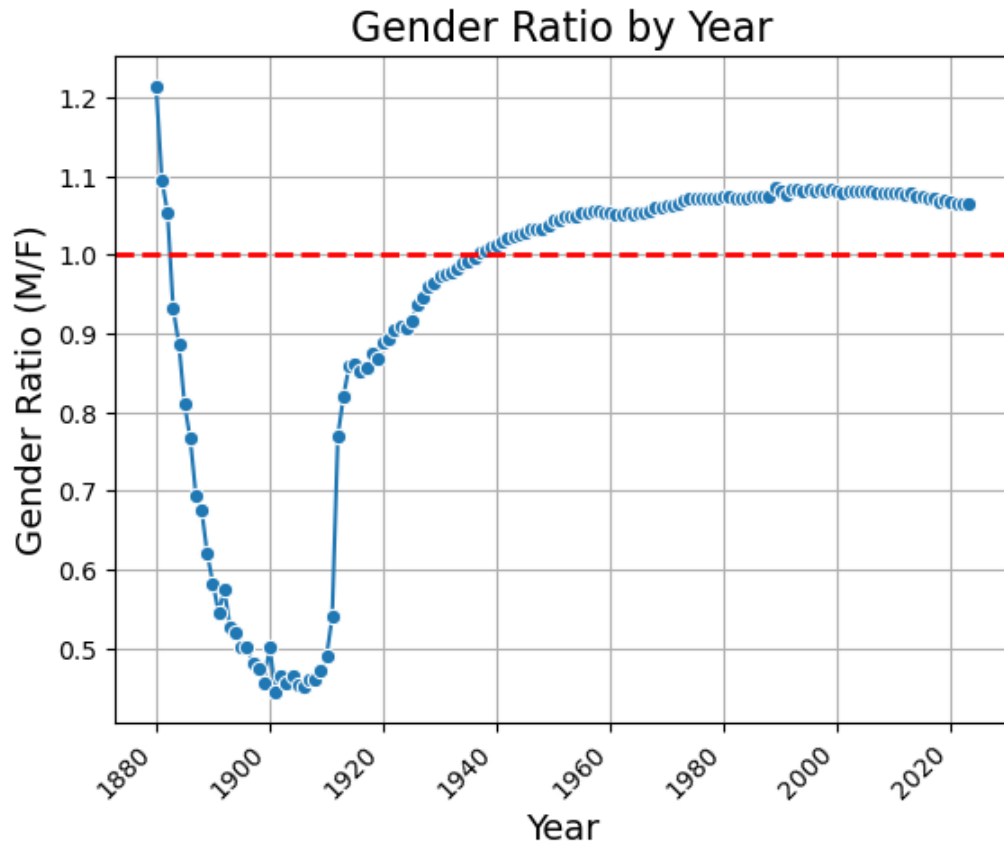
- **Business Expansion:** Growth in businesses and industries.
- **Progressive Reforms:** Social and political reforms aimed at addressing the inequalities and corrupt practices of the time.

## 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



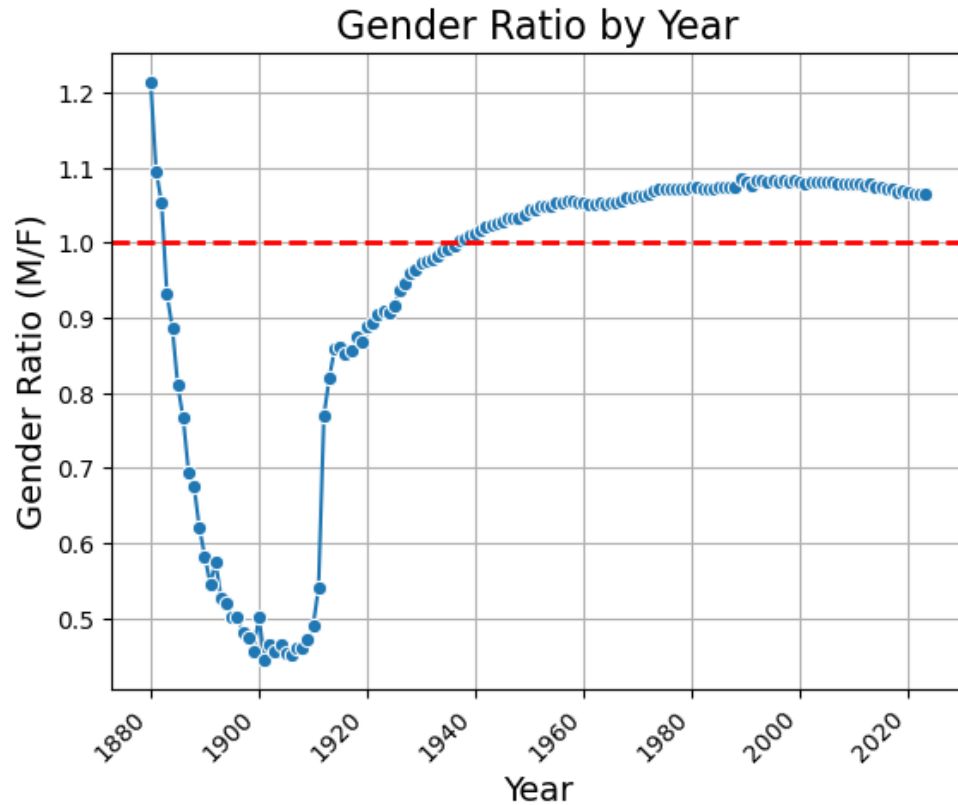
## Immigration (1880-1930):

- **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



## 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:** The relationship between societal stressors and gender ratios is influenced by multiple factors, requiring further empirical evidence.

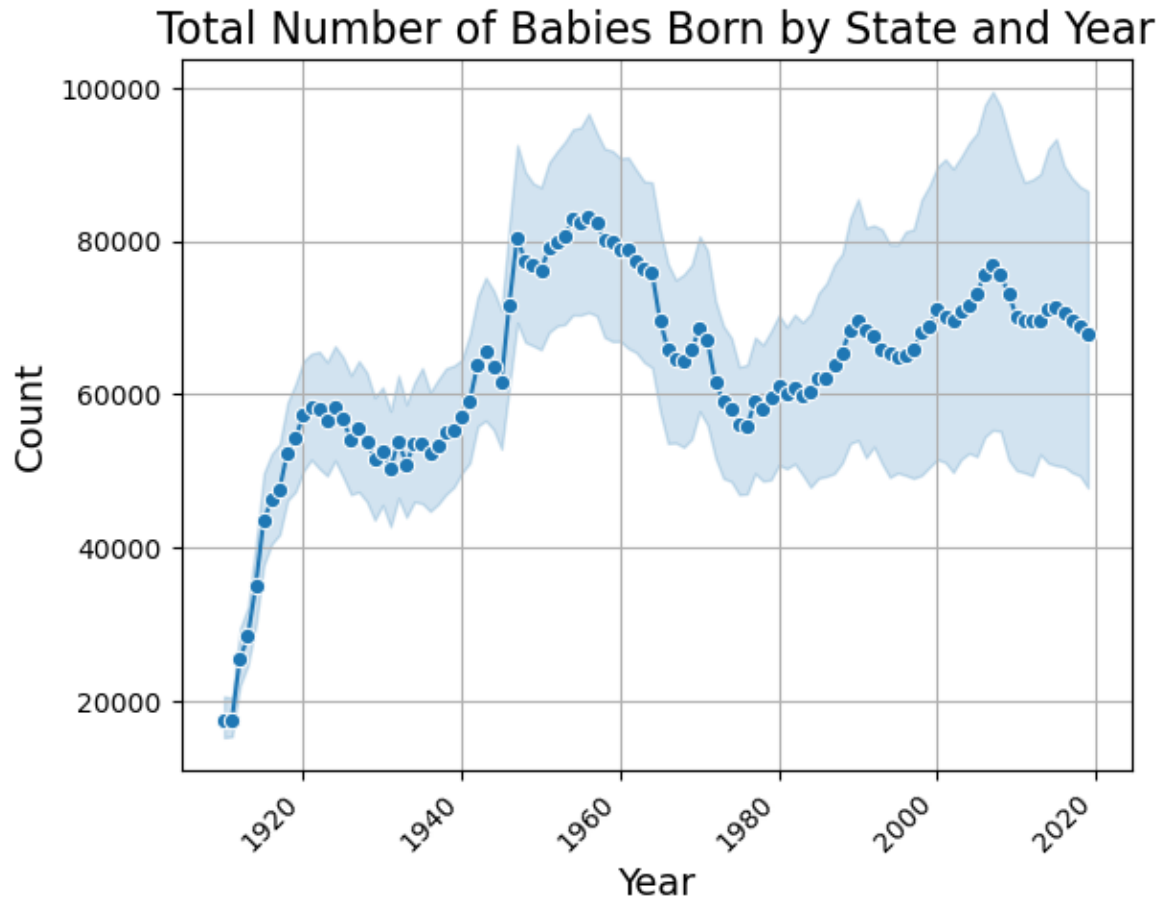
# Dataset2: State-Level Baby Name Data (1910-2021)

- 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*



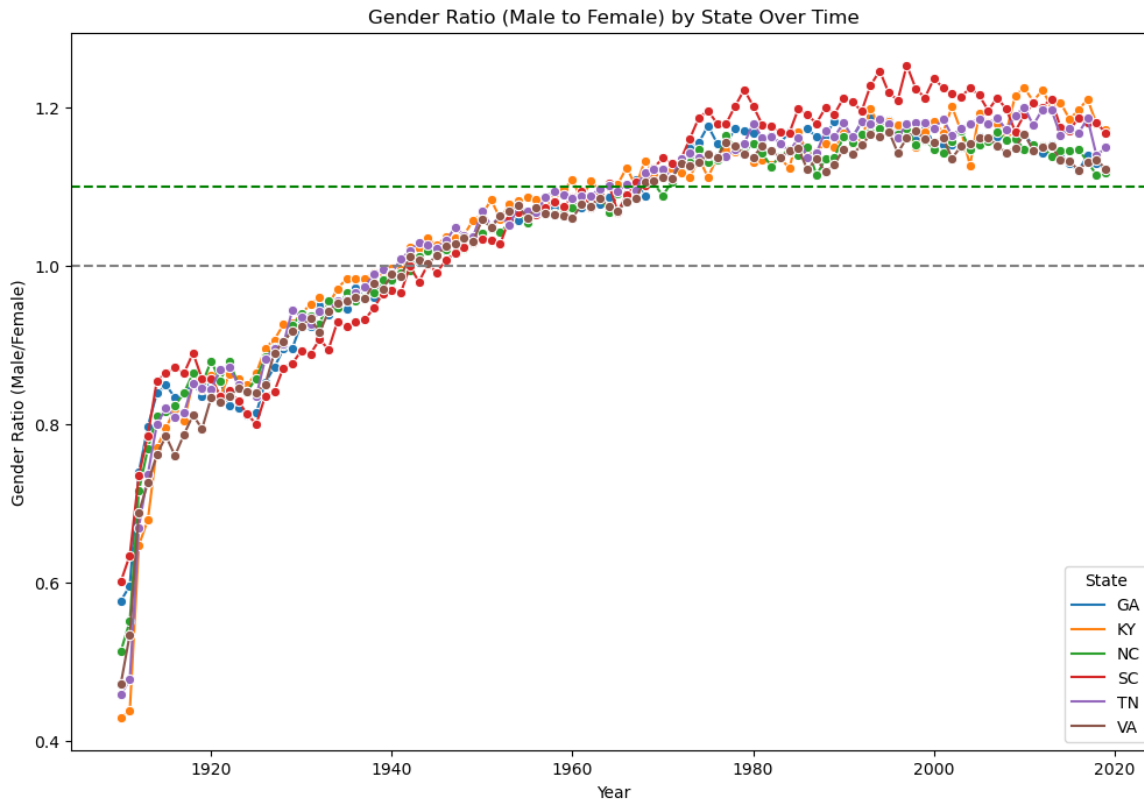
<https://www.newscientist.com/article/2416169-babies-in-bilingual-homes-have-distinct-brain-patterns-at-4-months-old/>

# Regional Analysis of Baby Birth Trends (1910-2021)



- **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 (1910-2021)



## 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.

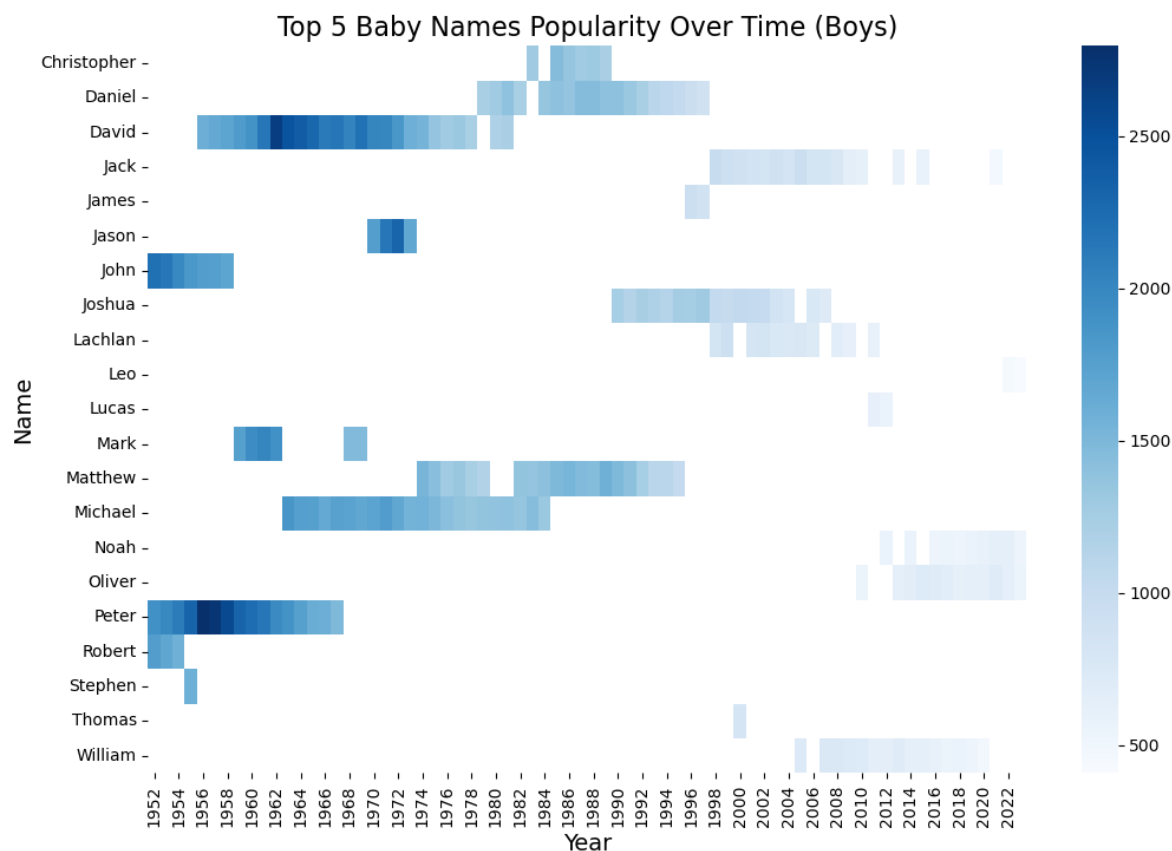
# Dataset3: Popular Baby Names (Top 100, 1952-2023)

- 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*



<https://alexischateau.com/2016/06/13/what-to-do-with-your-mixed-race-child/>

# Trends in Top 5 Baby Boy Names Over Time



- **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.*



<https://www.artistshot.com/euphoria-lotus-infant-clothing-p-12389910>