Trending Baby Names Data Analysis

Project Overview

This report presents an analysis of baby names data obtained from Social Security card applications in the United States. The dataset covers baby names from 1880 to 2022, including information about the name, gender, and the number of times each name was used.

Streamlit App:

In addition to the data analysis, a user-friendly Streamlit web application has been developed as part of this project. This app allows users to interactively explore the dataset and discover the top 10 baby names based on the gender and year they select. The app provides an intuitive interface for visualizing naming trends and exploring popular names over time.

The Streamlit app enhances the accessibility of the analysis results and provides an engaging way for users to explore the rich history of baby names in the United States.

Data Dictionary

Column Name	Description	Data Type
Name	The name of the person	String
Sex	The gender associated with the name (Male or Female).	Categorical
Number	The number of times that name was used in Social Security card applications.	Integer
Year	The year associated with each data record	Integer

Data Preparation

- Data Retrieval and Cleaning
- The data was extracted from a ZIP file containing yearly data files.
- We combined data from all years into a single dataset, creating a comprehensive view of baby naming trends over time.

Exploratory Data Analysis

- Popular Names Over Time
- Gender-Based Analysis
- Variations in Name Spelling

Visualization

- Trend Visualizations

  We visualized the popularity of names using line plots, allowing us to observe trends over time. We also normalized the data by the total number of births each year to highlight relative popularity.

- Stacked Area Plot

  A stacked area plot was used to visualize the popularity of variations of the name "Claire" over time.

- Top Names

  We identified the top ten names for each gender and year, creating tables for easy reference.

- Unisex Names

  We explored unisex names by calculating gender ratios and identifying names that fit within a certain range. Common unisex names were visualized to show their popularity for both genders.

Conclusion

This analysis of baby names data has unveiled captivating trends and patterns in naming preferences over more than a century. These trends provide valuable insights into cultural and societal changes, influencing the choices parents make when naming their children. While the specific names used for illustration purposes have seen variations, the broader trends in baby naming reflect the dynamic nature of society's values and influences. This analysis serves as a testament to the richness of historical data and its ability to reveal the evolving tapestry of our culture through something as simple yet meaningful as a name.

Future Work

In the future, there is significant potential to expand the capabilities of the Streamlit app, making it an even more versatile tool for exploring baby naming trends. Enhancements may include advanced user interaction features and deeper sociocultural investigations. Additionally, as the dataset continues to receive updates, the app's role in tracking contemporary naming trends will remain crucial, providing ongoing insights into the evolving landscape of baby names in the United States.