Baby Name Trend Analysis

Diego Sanoja

2025-06-04

Introduction

The purpose of this document is to explain the steps taken to perform the analysis using SQL in the MySQL database that contains information of names that parents gave to their babies when they were born. Furthermore, Power BI was used to create a report to show the insights found by the queries of the analysis.

Project Objective

The objective of this project was to dig into the baby names data to produce interesting insights about the baby names over the years for a fictional website company. Then, create a Power BI report to visualize and share the insights discovered and produce new ones.

Changelog

This section will provide a list of the modifications done to the data. Note: for the columns and measures created, a PDF document which contains the DAX codes used to create them will be presented in a separate file.

Version 1.0.0 (04-06-2025)

- Perform exploratory data analysis in the 2 tables of the database. To do this, queries that returned different information were written:
  - For the Names table, queries were written to find:
    - The year range in the table.
    - The different values in the Gender and State columns.
    - The total number of different names for boys and girls.
    - The presence of null/missing values in the columns.
- For the Regions table, queries were written to find:
  - The different values in the State column to compare it with the ones of the same column in the Names table.
  - The missing state in this table (using a JOIN).
- Write queries to discover the following information in the given order:

  - The overall most popular girl and boy names along with a column that shows their popularity (using the ROW_NUMBER window function).
  - The change in popularity of the names from the year 1980 compared to the one those names had in 2009.
  - The 3 most popular names for boys and girls per year and then per decade along with the number of babies that received those names for each case.
  - The total number of babies born per region over the 3 given decades.
  - The 10 most popular names used by both boys and girls.
  - The most popular short and long names.
  - The total percentage of babies given the name Chris per state.

Version 1.1.0 (05-06-2025)

- Import 3 tables derived from the database using queries into a new Power BI Desktop file for further analysis. Those tables are the following:
  - Regions which contains the names of the states and their regions.
  - Babies Per Year which contains the following 5 columns: Year, Gender, Name, Total_babies, and Popularity.
  - Babies Per State which contains the following 5 columns: State, Gender, Name, Total_babies, and Popularity.
- Create 2 report pages in the Power BI file called 'Babies Per Year' and 'Babies Per Region'.
- Add a new table to store the calculated measures for the report.
- Add the following calculated measures inside the new table:
  - Total Babies to calculate the total number of babies born.
  - % of Babies to compute the % of babies when using filters.
  - Total Males to calculate the number of male babies.
  - Total Females to compute the number of female babies.
  - Total Different Names to calculate the number of different names given to the babies.
- Store the previous calculations in a folder called 'Babies Per Year'.
- Add the following 2 calculations to the 'Babies Per Year' folder:
  - Male Popularity to compute the popularity of a specific name per year for boys.
  - Female Popularity to calculate the popularity of a specific name per year for girls.
- Add the following visuals to the 'Babies Per Year' page:
  - Three slicers on the right end to filter the data by gender, year and name.
  - A card in the top to display the values of the first 5 calculated measures in the folder for this page.
  - A line chart in the middle left to display the total male and female babies born per year.
  - A matrix in the lower left end to show the number of names given to boys and girls and their popularity per year.

- Repeat all the steps done for calculations and visuals in the 'Babies Per Year' page to fill the 'Babies Per State' page while doing the following changes:
  - Use the 'Babies Per State' table columns for the calculated measures.
  - Change the name of the folder to the one of the new page.
  - Add V2 at the end of the new calculations.
  - Replace the line chart by a clustered column chart.
  - Change the Year column in the filter by the Region column of the Regions table. Similarly, replace the year column in the matrix and column chart by the State column of the 'Babies Per State' table.

Version 1.2.0 (06-06-2025)

- Use the DISTINCT function to create a new table called 'Names Length' that stores the different baby names.
- Add a column in the new table using the LEN function to show the number of characters in each name.
- Connect the new table to the Babies Per Year & State tables using the Name columns.
- Add a new slicer to filter by the length of the name in the 'Babies Per Year' page.

Data Insights

This section will give a list of the insights provided both by the analysis done in MySQL and the visuals created in the pages of the Power BI report.

MySQL

This subsection will show each of the insights obtained through each phase of the analysis done using MySQL.

Initial Exploratory Data Analysis

The objective of this stage was to learn more about the tables that would be used later for the analysis and check for errors or typos which could generate inaccurate results in the future stages of the project. In this stage, the following insights were found:

For the Names table:

- The year range starts from 1980 and ends in 2009.
- There are no typos in the State and Gender columns.
- There are no missing values in any column.
- There are a total of 14474 and 9730 different names for girls and boys respectively but the total of different names is 22240 (which is lower than the sum of the 2 previous numbers). Therefore, there are names in the table which are given to both boys and girls.
- There are 51 different states in the table.

For the Regions table:

- There are 50 different states in the table (one less than in the Names table).
- The missing state in this table is MI which should be located in the Midwest region.

Track Changes In Name Popularity

The goal of this part was to discover the following information: the most popular names for boys and for girls in the 3 decades of the database, the popularity change of those names over the years, and which names present in both the first and last year had the greatest change in popularity. Five queries were written to answer the previous questions and they revealed the following information:

- The most popular name for a boy was Michael with a total of 1376418 babies receiving this name.
- The most popular name for a girl was Jessica with a total of 863121 babies receiving this name.
- From 1980 to 1997, Jessica was among the 3 most used names for girls with at least 20000 babies receiving that name each year but after that, its popularity decreased more every year.
- The overall popularity of the name Michael for boys remained high since it was the most popular during the first 19 years and the second most popular for the next 10 years.
- The 5 names whose popularity increased the most over the 30 year period are Colton, Aidan, Rowan, and Skylar.
- The 5 names whose popularity decreased the most over the 30 year period are Rusty, Tonia, Cherie, Kerri and Tara.

Compare Popularity Across Decades

The objective of this stage was to discover the 3 most used names for boys and girls per year and per decade. For this, a different query was written for each case and its results were sorted to obtain each insight. The following insights were obtained from those queries:

- For females, some popular names besides Jessica (which was the most used name for girls in the 80s and 90s) over the 3 decades of the database were:
  - Jennifer (most popular from 1980 to 1984 and second most popular in the 80s).
  - Amanda (either as the second or third most used for 7 years during the 80s and third most popular in the decade).
  - Ashley (on the top 3 names every year from 1984 to 1997 and second most popular of the 90s).
  - Emily (on the top 3 names every year from 1994 to 2008 and the most popular in the 2000s).
- For males, some popular names besides Michael (which was the most used name for boys in the 80s and 90s and second in the 2000s) over the 3 decades of the dataset were:
  - Christopher (second most used every year from 1980 to 1994 and second most popular in those decades).
  - Matthew (either as the second or third most popular every year of the 80s and 90s and second most used in those decades).
  - Jacob (most popular name from 1999 to 2009).
  - Joshua (third most popular every year from 2002 to 2006 and third one in the respective decade).
- There have been more changes in popularity over the 3 most used names for girls than for boys in the 3 decades.

Compare Popularity Across Regions

The goal of this phase was to discover the 3 most popular names for boys and girls per region and the total number of babies born in them over the 3 decades. For this, a different query was written for each case and its results were sorted to obtain each insight. Those queries revealed the following information:

- The regions with more and less babies born were the South and New England with a total of 34219920 and 4269213 respectively.
- For females:
  - Jessica was the most popular name in every region except for the South.
  - Ashley was in the 3 most used names for every region except in New England.
  - Jennifer was the third most popular name in the Mid Atlantic and the South and the second one in the Pacific.
  - Sarah was the third most used name in the Midwest and the Mountains and the second one in New England.
- For males:
  - Michael was the most popular name in every region except for the South.
  - Christopher was in the 3 most used names for every region except in Midwest.
  - Matthew was the second most popular name in the New England, the Midwest and the Mid Atlantic.
  - Joshua was the third most popular name in the Midwest and the South, and the second one in the Mountains.

Explore Unique Names

The objective of this stage was to search for information to discover the 10 most popular androgynous names (names given to both boys and girls), the length of the shortest and longest names and their popularity. To find that, 3 different queries were written and they revealed the following insights:

- There were a total of 1964 androgynous names from which the 4 most used are the male names shown in the previous section.
- The shortest names in the database have only 2 letters and there are 69 different names with that many characters. From those names, the most popular is Ty with 29205 babies that received this name.
- The longest names in the database have 15 letters and there are 4 names with that many characters: Franciscojavier (the most popular with 52 babies receiving it), Ryanchristopher, Johnchristopher and Mariodelosangel.

Power BI

This subsection will show the additional insights obtained by visualizing the names of the dataset by year and region in the 2 pages of the Power BI report.

Babies Per Year

This report page was built to generate insights from the distribution of the baby names by year. The visuals and filters in this page revealed the following information:

- Over the 3 decades of information in the dataset, a total of 98.73 million babies were born and received 1 of 22240 different names.
- From the babies in the dataset, 52.87 million (53.55%) of them are males and the rest were females.
- Every year, more boys were born than girls. Also, the year with less boys had more babies than the year with more girls.
- The years when more and less babies were born for both genders were 1990 and 1980 respectively.
- The decades with more and less births were the 90s and 80s respectively with a difference of 0.97%.
- Every name in the dataset was used at least 5 times (shown by sorting the matrix by the number of babies).
- Almost 81% of the babies in the dataset (about 79.8 million of them) have names that contain between 5 and 8 characters.

Babies Per State

This report page was built to generate insights from the distribution of the baby names by state. The visuals and filters in this page revealed the following information:

- The states with less and more babies born (of both genders) are CA (California) and WY (Wyoming) with a total of 14.08 millions (7.39 of them are males) and about 128680 babies (about 74580 of them are males).
- From the 4.27 million of the babies born in the New England region, almost half of them were born in the state of MA (Massachusetts).
- For the Mid Atlantic region, the state with more babies is NY (New York) which had a total of 6.69 million babies (3.55 million are males).
- The state with more babies (California) is located in the Pacific region and it had about 80% of the babies born in the region.
- The second state with more babies is TX (Texas) and is located in the South region which has 16 states of the 51 in the dataset.