

American Baby Names

What makes a name timeless or trendy? In this project, we'll use data published by the U.S. Social Security Administration spanning over a hundred years to understand American baby name tastes

How have American baby name tastes changed since 1920? Which names have remained popular for over 100 years, and how do those names compare to more recent top baby names?

We'll be working with data provided by the United States Social Security Administration, which lists first names along with the number and sex of babies they were given to each year. The data only shows first names which were given to over 5,000 American babies in a given year. Our data spans 101 years, from 1920 through 2020.

Let's see our table!

column	type	meaning
year	int	year
first_name	varchar	first name
sex	varchar	sex of babies given first_name
num	int	number of babies of sex given first_name in that year

We come to the most interesting part, we will execute some SQL queries and solve some practical questions.

Classic American Names



We will find names that have been given to over 5,000 babies every year for the 101 years from 1920 through 2020; recall that names only show up in our dataset when at least 5,000 babies have been given that name in a year.

```
-- We will select first names and the total babies with that first_name
-- Group by first_name and filter for those names that appear in all 101
years
-- Order by the total number of babies with that first_name in descending
order
```

```
SELECT first_name, SUM(num)
FROM baby_names
GROUP BY first_name
HAVING COUNT(year) = 101
ORDER BY SUM(num) DESC;
```

we get 8 rows affected :

first_name	sum
James	4748138
John	4510721
William	3614424
David	3571498
Joseph	2361382
Thomas	2166802
Charles	2112352
Elizabeth	1436286

It turns out that there are 8 classic timeless names that have appeared consistently in the last 101 years. 7 of the names seem to be named for boys and **James** is the most common name for boys! **Elizabeth** is holding her own for the female names, too.

Is it a classic name, or is it trendy?

Now, let's broaden our understanding of the dataset by looking at all names. We'll attempt to capture the type of popularity that each name in the dataset enjoyed. Was the name classic and popular across many years or trendy for a few years? Let's find out.

```
-- Classify first names as 'Classic', 'Semi-classic', 'Semi-trendy', or 'Trendy'
-- Alias this column as popularity_type
-- Select first_name, the sum of babies who have ever had that name, and popularity_type
-- Order the results alphabetically by first_name

SELECT first_name, SUM(num),
       CASE WHEN COUNT(year) > 80 THEN 'Classic'
            WHEN COUNT(year) > 50 THEN 'Semi-classic'
            WHEN COUNT(year) > 20 THEN 'Semi-trendy'
            ELSE 'Trendy' END AS popularity_type
FROM baby_names
GROUP BY first_name
ORDER BY SUM(num);
```

We found 547 rows of results. The top 5 results are Classic-Male names. The most chosen Classic name is **James**, **Christopher** for semi-classic, **Jennifer** for semi-trendy and **Brittany** for trendy names.

first_name	sum	popularity_type
James	4748138	Classic
John	4510721	Classic
Robert	4495199	Classic
Michael	4278824	Classic
William	3614424	Classic

Top Ranked Female Names since 1920

Since we didn't get many traditionally female names in our classic American names search in the first query, let's filter our search to names which were given to female babies.

We can use this opportunity to practice window functions by assigning a rank to female names based on the number of babies that have ever been given that name. What are the top-ranked female names since 1920?

```
-- RANK names by the sum of babies who have ever had that name
(descending), aliasing as name_rank
-- Select name_rank, first_name, and the sum of babies who have ever had
that name
-- Filter the data for results where sex equals 'F'
-- Limit to ten results
```

```
SELECT RANK() OVER(ORDER BY SUM(num) desc) AS name_rank,
       first_name,
       SUM(num)
FROM baby_names
WHERE sex = 'F'
GROUP BY first_name
LIMIT 10
```

result :

name_rank	first_name	sum
1	Mary	3215850
2	Patricia	1479802
3	Elizabeth	1436286
4	Jennifer	1404743
5	Linda	1361021
6	Barbara	1343901
7	Susan	1025728
8	Jessica	994210
9	Lisa	920119
10	Betty	893396

It turns out that **Mary, Patricia and Elizabeth** is the top 3 first name for 1920 - 2021 for girls name in America.

Help! My friend needs our help to pick a baby name!

Perhaps a friend has heard of our work analyzing baby names and would like help choosing a name for her baby, a girl. She doesn't like any of the top-ranked names we found before.

She's set on a traditionally female name ending in the letter 'a' since she's heard that vowels in baby names are trendy. She's also looking for a name that has been popular in the years since 2015.

Let's see what we can do to find some options for this friend!

```
-- Select only the first_name column
-- Filter for results where sex is 'F', year is greater than 2015, and
first_name ends in 'a'
-- Group by first_name and order by the total number of babies given that
first_name
```

```
SELECT first_name, SUM(num)
FROM baby_names
WHERE sex = 'F' AND year > 2015 AND first_name LIKE '%a'
GROUP BY first_name
ORDER BY SUM(num) DESC
LIMIT 5
```

Well, we can see that **Olivia**, **Emma** and **Ava** is the most used name for girls. Let's give the result to my friend, and she will choose the name.

first_name	sum
Olivia	92178
Emma	90865
Ava	74856
Sophia	71760
Isabella	69977

The Expansion of Olivia

Surprisingly, we can see that Olivia is the most popular female name ending in 'A' since 2015. When did the name Olivia become so popular? Let's explore the rise of the name Olivia

```
-- Select year, first_name, num of Olivias in that year, and
cumulative_olivias
-- Sum the cumulative babies who have been named Olivia up to that year;
alias as cumulative_olivias
-- Filter so that only data for the name Olivia is returned.
-- Order by year from the earliest year to most recent
```

```
SELECT year, first_name, num,
       SUM(num) OVER (ORDER BY year) AS cumulative_olivias
FROM baby_names
WHERE first_name = 'Olivia'
ORDER BY year;
```

Results:

year	first_name	num	cumulative_olivias
1991	Olivia	5601	5601
1992	Olivia	5809	11410
1993	Olivia	6340	17750
...			
2013	Olivia	18439	297407
2014	Olivia	19823	317230
2015	Olivia	19710	336940
...			
2019	Olivia	18508	411583
2020	Olivia	17535	429118

Olivia first appeared in 1991 with a total of 5,601. The largest increase occurred in 2014, where there were 19,823 additions of Olivia in that year. Over time, **Olivia has been used 429,118 times in 2020**. What a meteoric rise!

Many Males with The Same Name

We saw in the first query (**Classic American Names**) that there are **seven** traditionally male names given to at least 5,000 babies every single year in our 101-year dataset! Those names are classics, but showing up in the dataset every year doesn't necessarily mean that the timeless names were the most popular. Let's explore popular male names a little further.

In the next two queries, we will build up to listing every year along with the most popular male name in that year. This presents a common problem: how do we find the greatest X in a group? Or, in the context of this problem, how do we find the male name given to the highest number of babies in a year?

Let's use a subquery. We can first write a query that selects the year and the maximum num of babies given any single male name in that year. For example, in 1989, the male name given to the highest number of babies was given to 65,339 babies. We'll write this query in this task. In the next task, we can use the code from this task as a subquery to look up the first_name that was given to 65,339 babies in 1989... as well as the top male first name for all other years!

```
-- Select year and maximum number of babies given any one male name in that
year, aliased as max_num
-- Filter the data to include only results where sex equals 'M'
-- We will use this query as sub query, later
```

```
SELECT year,
       MAX(num) as max_num
FROM baby_names
WHERE sex = 'M'
GROUP BY year
```

There are 101 rows affected, and here is the top 5 rows :

year	max_num
1920	56914
1921	58215
1922	57280
1923	57469
1924	60801

Top male names over the years 🧑

Let's find the most used names each year. Use sub-query and Inner Join to get the top male names over the year

```
-- Select year, first_name given to the largest number of male babies, and
-- num of babies given that name
-- Join baby_names to the code in the last task as a subquery
-- Order results by year descending
```

```
SELECT b.year, b.first_name, b.num
FROM baby_names AS b
INNER JOIN (
    SELECT year, MAX(num) as max_num
    FROM baby_names
    WHERE sex = 'M'
    GROUP BY year) AS subquery
ON subquery.year = b.year
AND subquery.max_num = b.num
ORDER BY year DESC;
```

there are 101 rows affected after this inner join, here are the top 5 rows:

year	first_name	num
2020	Liam	19659
2019	Liam	20555
2018	Liam	19924
2017	Liam	18824
2016	Noah	19154

We can see that in 2020, Liam is the most used by Americans. it was used 19,659 times.

The most years at number one 🏆

Noah and Liam have ruled the roost in the last few years, but if we scroll down in the results, it looks like Michael and Jacob have also spent a good number of years as the top name! Which name has been number one for the largest number of years? Let's use a common table expression to find out.

```
-- Select first_name and a count of years it was the top name in the last
task; alias as count_top_name
-- Use the code from the previous task as a common table expression
-- Group by first_name and order by count_top_name descending
```

```
WITH top_male_names AS (
    SELECT b.year, b.first_name, b.num
    FROM baby_names AS b
    INNER JOIN (
        SELECT year, MAX(num) num
        FROM baby_names
        WHERE sex = 'M'
        GROUP BY year) AS subquery
    ON subquery.year = b.year
    AND subquery.num = b.num
    ORDER BY YEAR DESC
)
SELECT first_name, COUNT(first_name) as count_top_name
FROM top_male_names
GROUP BY first_name
ORDER BY COUNT(first_name) DESC;
```

first_name	count_top_name
Michael	44
Robert	17
Jacob	14
James	13
Noah	4
John	4
Liam	4
David	1

What an unexpected result! We found out that Michael is America's favorite male name for 44 years.