



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? These are considerations for many new parents, but the skills you'll practice while answering these queries are broadly applicable. After all, understanding trends and popularity is important for many businesses, too!

You'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 in each year. For processing speed purposes, the dataset is limited to first names which were given to over 5,000 American babies in a given year. The data spans 101 years, from 1920 through 2020.

The Data

baby_names				
column	type		description	
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

Projects Data DataFrame as usa\_baby\_names

```
-- Run this code to view the data in baby_names
SELECT *
FROM baby_names
LIMIT 50;
```

index	...	↑↓	year	...	↑↓	first_name	...	↑↓	sex
		0			1920	Mary			F
		1			1920	Dorothy			F
		2			1920	Helen			F
		3			1920	Margaret			F
		4			1920	Ruth			F
		5			1920	Mildred			F
		6			1920	Virginia			F
		7			1920	Elizabeth			F
		8			1920	Frances			F
		9			1920	Anna			F
		10			1920	Betty			F
		11			1920	Evelyn			F
		12			1920	Marie			F
		13			1920	Doris			F
		14			1920	Alice			F
		15			1920	Florence			F

Rows: 50

Expand

Projects Data DataFrame as name\_types

```
-- Use this table for the answer to question 1:
-- List the overall top five names in alphabetical order and find out if each name is "Classic" or "Trendy."
SELECT
    first_name,
    SUM(num) AS sum,
    CASE
        WHEN (COUNT(*) >= 50) THEN 'Classic'
        ELSE 'Trendy'
    END AS popularity_type
FROM baby_names
GROUP BY first_name
ORDER by first_name ASC
LIMIT 5
```

index	...	↑↓	first_name	...	↑↓	sum	...	↑↓	popularity_
		0	Aaliyah				15870		Trendy
		1	Aaron				530592		Classic
		2	Abigail				338485		Trendy
		3	Adam				497293		Trendy
		4	Addison				107433		Trendy

Rows: 5

Expand

```
-- Use this table for the answer to question 2:
-- What were the top 20 male names overall, and how did the name Paul rank?
```

```
SELECT
  RANK() OVER (ORDER BY SUM(num) DESC) AS name_rank,
  first_name,
  SUM(num) AS sum
FROM baby_names
WHERE sex = 'M'
GROUP BY first_name
LIMIT 20
```

index	...	↑↓	name_rank	...	↑↓	first_name
		0			1	James
		1			2	John
		2			3	Robert
		3			4	Michael
		4			5	William
		5			6	David
		6			7	Richard
		7			8	Joseph
		8			9	Thomas
		9			10	Charles
		10			11	Christopher
		11			12	Daniel
		12			13	Matthew
		13			14	Anthony
		14			15	Donald
		15			16	Mark

Rows: 20

[Expand](#)

```
-- Use this table for the answer to question 3:
-- Which female names appeared in both 1920 and 2020?
```

```
WITH name_1920 AS (
  SELECT *
  FROM baby_names
  WHERE
    sex = 'F' AND
    year = 1920
),
name_2020 AS (
  SELECT *
  FROM baby_names
  WHERE
    sex = 'F' AND
    year = 2020
)
SELECT
  name_1920.first_name,
  name_1920.num + name_2020.num AS total_occurrences
FROM name_1920
JOIN name_2020
  ON name_1920.first_name = name_2020.first_name
```

...	↑↓	fi...	...	↑↓	total_occurre...	...	↑↓
0		Emma			20818		
1		Evelyn			23283		
2		Elizabeth			23125		
3		Eleanor			14832		
4		Grace			12741		
5		Hazel			12765		

Rows: 6

[Expand](#)