

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 5;
```

index	...	↑↓	year	...	↑↓	first_name	...	↑↓	sex	...	↑↓	num	...
		0			1920	Mary			F				70
		1			1920	Dorothy			F				36
		2			1920	Helen			F				35
		3			1920	Margaret			F				27
		4			1920	Ruth			F				26

Rows: 5 

 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(year) >= 50 THEN 'Classic'
    ELSE 'Trendy' END AS popularity_type
FROM baby_names
GROUP BY first_name
ORDER BY first_name
LIMIT 5;
```

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

Rows: 5 [↓](#)

 Projects Data DataFrame as top_20

-- 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	...	↑↓	sum	...	↑↓
		0			1	James			4748		
		1			2	John			4510		
		2			3	Robert			4495		
		3			4	Michael			4278		
		4			5	William			3614		
		5			6	David			3571		
		6			7	Richard			2414		
		7			8	Joseph			2361		
		8			9	Thomas			2166		
		9			10	Charles			2112		
		10			11	Christopher			2012		
		11			12	Daniel			1824		
		12			13	Matthew			1567		
		13			14	Anthony			1344		
		14			15	Donald			1280		
		15			16	Mark			1265		

Rows: 20 [↓](#)

 Projects Data DataFrame as a_names

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

```
SELECT
  first_name,
  SUM(num) AS total_occurrences
FROM baby_names
WHERE sex = 'F' AND first_name IN
  (SELECT a_19.first_name
   FROM baby_names AS a_19
   INNER JOIN (SELECT a_20.first_name
               FROM baby_names AS a_20
```

```
WHERE a_20.year = 2020) AS sub
USING(first_name)
WHERE a_19.year = 1920)
GROUP BY first_name;
```

...	↑↓	fi...	...	↑↓	total_occurre...	...	↑↓
0		Grace			254573		
1		Elizabeth			1436286		
2		Eleanor			119863		
3		Evelyn			310824		
4		Hazel			66103		
5		Emma			448087		