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;
```

	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 rows

Projects Data DataFrame as `name_types`

```
-- Use this table for the answer to question 1:
```

```sql
-- List the overall top five names in alphabetical order and find out if each name is
"Classic" or "Trendy."
-- Select first_name, the sum of babies who have ever had that name, and popularity_type
SELECT first_name, SUM(num),
-- Classify first names as 'Classic' or 'Trendy'
    CASE WHEN COUNT(year) > 50 THEN 'Classic'
        ELSE 'Trendy' END AS popularity_type
FROM baby_names
-- Group by first_name to use aggregate functions
GROUP BY first_name
-- Order the results alphabetically by first_name
ORDER BY first_name
-- Limit to the first 5 names
LIMIT 5;
```

	first_name	sum	popul
0	Aaliyah	15870	Trend
1	Aaron	530592	Class
2	Abigail	338485	Trend
3	Adam	497293	Trend
4	Addison	107433	Trend

5 rows

Projects Data DataFrame as top_20

```sql
-- Use this table for the answer to question 2:
-- What were the top 20 male names overall, and how did the name Paul rank?
-- RANK names by the sum of babies who have ever had that name (descending), aliasing as
name_rank
SELECT
    RANK() OVER(ORDER BY SUM(num) DESC) AS name_rank,
    first_name, SUM(num)
FROM baby_names
-- Filter the data for results where sex equals 'M'
WHERE sex = 'M'
-- Group by first name, oder by rank, and limit to the top 20
GROUP BY first_name
ORDER BY name_rank
LIMIT 20;
```

	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

Projects Data DataFrame as a

```sql
-- Use this table for the answer to question 3:
-- Which female names appeared in both 1920 and 2020?
-- Select first name and total occurrences
SELECT a.first_name, (a.num + b.num) AS total_occurrences
FROM baby_names a
JOIN baby_names b
-- Join on first name
ON a.first_name = b.first_name
-- Filter for the years 1920 and 2020 and sex equals 'F'
WHERE a.year = 1920 AND a.sex = 'F'
AND b.year = 2020 AND b.sex = 'F';
```

	first_name	total_occurrences
0	Emma	
1	Evelyn	
2	Elizabeth	
3	Eleanor	
4	Grace	
5	Hazel	

6 rows