PROJECT: EXPLORING TRENDS IN AMERICAN BABY NAMES

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

```
-- Run this code to view the data in baby_names
SELECT *
FROM baby_names
LIMIT 5;
```

		fi...		
0	1920	Mary	F	70982
1	1920	Dorothy	F	36643
2	1920	Helen	F	35097
3	1920	Margaret	F	27994
4	1920	Ruth	F	26101

Rows: 5

Projects Data DataFrame as

```
-- 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),
    CASE WHEN COUNT(*) > 10 THEN 'Over Ten'
        ELSE 'Not Over Ten' END AS popularity_type
FROM baby_names
GROUP BY first_name
ORDER BY first_name
LIMIT 5;
```

	fi...		popularity...
0	Aaliyah	15870	Not Over Ten
1	Aaron	530592	Over Ten
2	Abigail	338485	Over Ten
3	Adam	497293	Over Ten
4	Addison	107433	Over Ten

Rows: 5

Projects Data DataFrame as

```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?

SELECT
    RANK() OVER(ORDER BY first_name DESC) AS name_rank,
    first_name,
    SUM(num)
FROM baby_names
WHERE sex = 'M' -- filter for male name
GROUP BY first_name
ORDER BY name_rank
LIMIT 20;
```

	n...	fi...	...
0	1	Zachary	483955
1	2	Xavier	51892
2	3	Wyatt	128168
3	4	Willie	274564
4	5	William	36144...
5	6	Wayne	211347
6	7	Warren	13290
7	8	Walter	378194
8	9	Vincent	23419
9	10	Tyler	548624
10	11	Troy	82294
11	12	Tristan	27212
12	13	Trevor	76138
13	14	Travis	218731
14	15	Tony	96417
15	16	Tom	5061
16	17	Todd	207137

Rows: 20

Projects Data DataFrame as

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

SELECT DISTINCT
    b1.first_name AS first_name,
    (b1.count + b2.count) AS total_occurrences
FROM baby_names AS b1
INNER JOIN baby_names AS b2
    ON b1.first_name = b2.first_name  -- Match names, not years
WHERE
    b1.sex = 'F'  -- Filter for female names
    AND b2.sex = 'F'  -- Filter for female names
    AND b1.year = 1920  -- Names from 1920
    AND b2.year = 2020  -- Names from 2020
GROUP BY b1.first_name;
```

	fi...	total_occurre...
0	Eleanor	2
1	Elizabeth	2
2	Emma	2
3	Evelyn	2
4	Grace	2
5	Hazel	2

Rows: 6