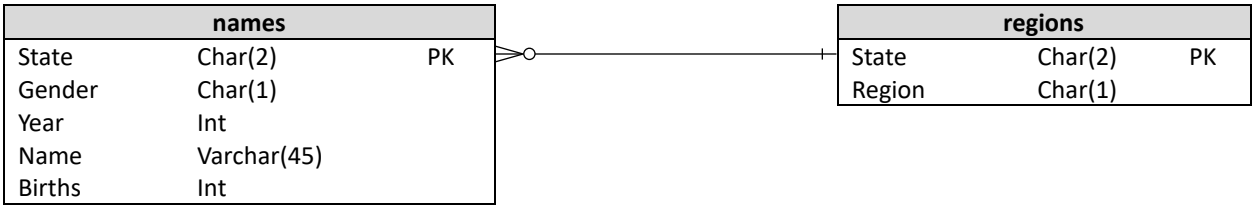
Project: U.S. Baby Names Analysis (1980–2019)

Tool: SQL

1. **Project Overview**
    - Dataset: U.S. Social Security Administration
    - Sector: Public Sector / Demographic Analytics
    - Size: Over 1.9 million rows
    - Tables: names (State, Gender, Year, Name, Births), regions (State, Region)
    - Time Range: 1980 to 2019
    - Objective: Analyze naming patterns and diversity over time and geography
    - Table relationship Diagram:



2. **Problem Statement:**
    - How have baby name preferences evolved across time, gender, and regions?
    - Can we extract trends to support product targeting or cultural understanding?

3. **Business Use Case:**
    - Identify top names by gender and state to tailor localized marketing
    - Track cultural shifts through naming patterns
    - Study generational preferences and diversity in naming conventions agencies studying population trends.

4. **Analysis Goals**
    - Identify most and least popular names
    - Discover naming trends across years
    - Compare naming diversity and uniqueness

5. **Metrics**
    - Name frequency
    - Name rank by year/gender
    - Name ranking movement
    - Name diversity (number of unique names per year)

6. **SQL to Insight**

    6.1. Most Popular Names (Overall)

```sql
-- 7.1.1 Popular girl name:
SELECT Name, SUM(Births) AS num_babies
FROM names
WHERE Gender = 'F'
GROUP BY Name
ORDER BY num_babies DESC
LIMIT 1;
```

Name	num_babies
Jessica	863121

```sql
-- 7.1.2. Popular boy name:
SELECT Name, SUM(Births) AS num_babies
FROM names
WHERE Gender = 'M'
GROUP BY Name
ORDER BY num_babies DESC
LIMIT 1;
```

Name	num_babies
Michael	1376418

=> Michael (M) and Jessica (F) were the most popular names across 1980–2019

6.2. Fastest-Rising Names

```sql
-- 7.2. the biggest jumps in the names' rankings over the years
WITH    Sub_Ranking_Table AS
        (WITH   Ranking_1980 AS (SELECT Name,
                row_number () OVER (ORDER BY SUM(Births) DESC) AS Ranking_in_1980
FROM names
WHERE Year = 1980
GROUP BY Name),
        Ranking_2009 AS (SELECT Name,
                row_number () OVER (ORDER BY SUM(Births) DESC) AS Ranking_in_2009
FROM names
WHERE Year = 2009
GROUP BY Name)
SELECT Ranking_1980.Name, Ranking_in_1980, Ranking_in_2009
FROM Ranking_1980
JOIN Ranking_2009 ON Ranking_1980.Name = Ranking_2009.Name)
SELECT Name, Ranking_in_1980, Ranking_in_2009, CAST(Ranking_in_1980 AS SIGNED) - CAST(Ranking_in_2009 AS SIGNED) AS Jumps
FROM Sub_Ranking_Table
GROUP BY Name
ORDER BY Jumps DESC
LIMIT 10;
```

Name	Ranking_in_1980	Ranking_in_2009	Jumps
Skylar	5781	313	5468
Rylan	5615	320	5295
Kyler	5780	511	5269
Lexi	5788	537	5251
Aliyah	5469	351	5118
Colton	5264	149	5115
Aidan	5166	109	5057
Griffin	5292	439	4853
Norah	5513	672	4841
Kenzie	5691	857	4834

=> Names like Skylar, Rylan, and Kyler showed dramatic ranking increases, despite not reaching top popularity overall

6.3. Most Popular Names by Decade

```sql
-- 7.3. For each decade, return the 3 most popular girl names and 3 most popular boy names
SELECT
    Decade,
    MAX(CASE WHEN Gender = 'M' AND popularity = 1 THEN Name END) AS Male_1st,
    MAX(CASE WHEN Gender = 'M' AND popularity = 2 THEN Name END) AS Male_2nd,
    MAX(CASE WHEN Gender = 'M' AND popularity = 3 THEN Name END) AS Male_3rd,
    MAX(CASE WHEN Gender = 'F' AND popularity = 1 THEN Name END) AS Female_1st,
    MAX(CASE WHEN Gender = 'F' AND popularity = 2 THEN Name END) AS Female_2nd,
    MAX(CASE WHEN Gender = 'F' AND popularity = 3 THEN Name END) AS Female_3rd
FROM (
    WITH Decade_Table AS (SELECT *,
        CASE
            WHEN Year < 1990 THEN '1980s'
            WHEN Year < 2000 THEN '1990s'
            ELSE '2000s'
        END AS Decade
    From Names)
    SELECT
        Decade,
        Name,
        Gender,
        ROW_NUMBER() OVER (PARTITION BY Decade, Gender ORDER BY SUM(Births) DESC) AS popularity
    FROM Decade_Table
    GROUP BY Decade, Name, Gender
) AS Ranking_Table
WHERE popularity <= 3
GROUP BY Decade
ORDER BY Decade ASC;
```

Decade	Male_1st	Male_2nd	Male_3rd	Female_1st	Female_2nd	Female_3rd
1980s	Michael	Christopher	Matthew	Jessica	Jennifer	Amanda
1990s	Michael	Christopher	Matthew	Jessica	Ashley	Emily
2000s	Jacob	Michael	Joshua	Emily	Madison	Emma

1980s: Michael, Christopher, Jessica
1990s: Jacob, Joshua, Emily
2000s: Madison, Emma

6.4. Regional Name Preferences

```
WITH adjusted_regions AS (
        SELECT  State,
                CASE WHEN Region = 'New England' THEN 'New_England' ELSE Region END AS adjusted_region
        FROM regions
        UNION
        SELECT  'MI' AS State, 'Midwest' AS REgion
        FROM regions),
Total_Table AS (SELECT          t1.State,
                                t2.adjusted_region,
                                t1.Year,
                                t1.Name,
                                t1.Gender,
                                t1.Births
From names t1
JOIN adjusted_regions t2
        ON t1.State = t2.State),
Regions_Birth_popularity AS (SELECT     adjusted_region AS Region,
        Name,
        Gender,
        SUM(births) AS Total_births,
        ROW_NUMBER() OVER (PARTITION BY adjusted_region, Gender ORDER BY SUM(births) DESC) AS Regions_popularity
FROM Total_Table
GROUP BY Region, Gender, Name
ORDER BY Total_births DESC)
SELECT  Region,
        MAX(CASE WHEN Gender ='M' AND Regions_popularity ='1' THEN Name END) AS 1st_Regions_Male_Name,
        MAX(CASE WHEN Gender ='M' AND Regions_popularity ='2' THEN Name END) AS 2nd_Regions_Male_Name,
        MAX(CASE WHEN Gender ='M' AND Regions_popularity ='3' THEN Name END) AS 3rd_Regions_Male_Name,
        MAX(CASE WHEN Gender ='F' AND Regions_popularity ='1' THEN Name END) AS 1st_Regions_Female_Name,
        MAX(CASE WHEN Gender ='F' AND Regions_popularity ='2' THEN Name END) AS 2nd_Regions_Female_Name,
        MAX(CASE WHEN Gender ='F' AND Regions_popularity ='3' THEN Name END) AS 3rd_Regions_Female_Name
FROM Regions_Birth_popularity
WHERE Regions_popularity <=3
GROUP BY Region;
```

Region	1st_Regi_M_Name	2nd_Regi_M_Name	3rd_Regi_M_Name	1st_Regi_F_Name	2nd_Regi_F_Name	3rd_Regi_F_Name
South	Christopher	Michael	Joshua	Ashley	Jessica	Jennifer
Midwest	Michael	Matthew	Joshua	Jessica	Ashley	Sarah
Mid_Atlantic	Michael	Matthew	Christopher	Jessica	Ashley	Jennifer
Pacific	Michael	Christopher	Daniel	Jessica	Jennifer	Ashley
New_England	Michael	Matthew	Christopher	Jessica	Sarah	Emily
Mountain	Michael	Joshua	Christopher	Jessica	Ashley	Sarah

=> Michael and Jessica were consistently top names across all U.S. regions, showing strong nationwide influence

6.5. Name Diversity Trends

```
-- 7.5 To find unique name number in dataset
SELECT year, COUNT(DISTINCT name) AS unique_names
FROM names
GROUP BY year
ORDER BY year;
```

year	unique_names	year	unique_names	year	unique_names
1980	5791	1990	7825	2000	8703
1981	5869	1991	7847	2001	8752
1982	5971	1992	7918	2002	8869
1983	5908	1993	7994	2003	9081
1984	6009	1994	7982	2004	9248
1985	6187	1995	7898	2005	9435
1986	6388	1996	7926	2006	9752
1987	6555	1997	7963	2007	10017
1988	6878	1998	8130	2008	9962
1989	7373	1999	8293	2009	9999

=> Name uniqueness increased yearly, peaking in 2007 with over 10,000 unique names, a shift toward personalized or unique baby names

7. **Recommendations**
    - Use name data to personalize baby-focused product campaigns
    - Monitor long-term trends for branding/media positioning
    - Leverage regional naming data for localized outreach
8. **The next steps:**
    - Build interactive dashboards by region/gender
    - Join with ethnicity/income/education data for segmentation
    - Apply clustering to identify naming behavior patterns