

# Data Analysis - Baby Names

## Objection 1: Track changes in name popularity

### Task 1

1. Find the overall most popular girl and boy names

| Gender | Name    | num_babies | name_rank |
|--------|---------|------------|-----------|
| M      | Michael | 1376418    | 1         |
| F      | Jessica | 863121     | 1         |

2. Show how they have changed in popularity rankings over the years

| Year | Name    | popularity | Year | Name    | popularity |
|------|---------|------------|------|---------|------------|
| 1980 | Michael | 1          | 2009 | Michael | 3          |
| 1981 | Michael | 1          | 2008 | Michael | 2          |
| 1982 | Michael | 1          | 2007 | Michael | 2          |
| 1983 | Michael | 1          | 2006 | Michael | 2          |
| 1984 | Michael | 1          | 2005 | Michael | 2          |
| 1985 | Michael | 1          | 2004 | Michael | 2          |
| 1986 | Michael | 1          | 2003 | Michael | 2          |
| 1987 | Michael | 1          | 2002 | Michael | 2          |
| 1988 | Michael | 1          | 2001 | Michael | 2          |
| 1989 | Michael | 1          | 2000 | Michael | 2          |
| 1990 | Michael | 1          | 1999 | Michael | 2          |
| 1991 | Michael | 1          |      |         |            |
| 1992 | Michael | 1          |      |         |            |
| 1993 | Michael | 1          |      |         |            |
| 1994 | Michael | 1          |      |         |            |
| 1995 | Michael | 1          |      |         |            |
| 1996 | Michael | 1          |      |         |            |
| 1997 | Michael | 1          |      |         |            |
| 1998 | Michael | 1          |      |         |            |

Michael was the number boy name in popularity in 1980 and stayed that way all through 1998. After that, it was the second or third most popular boy name throughout 2009.

| Year | Name    | popularity | Year | Name    | popularity |
|------|---------|------------|------|---------|------------|
| 1980 | Jessica | 3          | 2009 | Jessica | 78         |
| 1981 | Jessica | 2          | 2008 | Jessica | 60         |
| 1982 | Jessica | 2          | 2007 | Jessica | 42         |
| 1983 | Jessica | 2          | 2006 | Jessica | 32         |
| 1984 | Jessica | 2          | 2005 | Jessica | 27         |
| 1985 | Jessica | 1          | 2004 | Jessica | 21         |
| 1986 | Jessica | 1          | 2003 | Jessica | 18         |
| 1987 | Jessica | 1          | 2002 | Jessica | 16         |
| 1988 | Jessica | 1          | 2001 | Jessica | 11         |
| 1989 | Jessica | 1          | 2000 | Jessica | 8          |
| 1990 | Jessica | 1          | 1999 | Jessica | 9          |
| 1991 | Jessica | 2          | 1998 | Jessica | 8          |
| 1992 | Jessica | 2          |      |         |            |
| 1993 | Jessica | 1          |      |         |            |
| 1994 | Jessica | 1          |      |         |            |
| 1995 | Jessica | 1          |      |         |            |
| 1996 | Jessica | 2          |      |         |            |
| 1997 | Jessica | 2          |      |         |            |

Jessica was top three in popularity for girl names from 1980 till 1997. However, it dropped as low as 78th most popular by 2009.

## **Task 2**

1. Find the names with the biggest jumps in popularity from the first year to the last year

| Year | Name   | popularity | Year | Name   | popularity | diff  |
|------|--------|------------|------|--------|------------|-------|
| 1980 | Aidan  | 5716       | 2009 | Aidan  | 109        | -5607 |
| 1980 | Skylar | 5523       | 2009 | Skylar | 313        | -5210 |
| 1980 | Macy   | 5640       | 2009 | Macy   | 571        | -5069 |
| 1980 | Kiara  | 5613       | 2009 | Kiara  | 568        | -5045 |
| 1980 | Rylan  | 5357       | 2009 | Rylan  | 320        | -5037 |

I sorted the popularity of names in 1980 (the first year of the dataset) and 2009 (the last year in the dataset). The right hand column shows the difference in change of popularity.

Aiden had the biggest jump going from being ranked 5716 in popularity in 1980 to 109 in 2009. Skylar, Macy, Kiara and Rylan followed in order for the next biggest jumps over the years.

## Objection 2: Compare popularity across decades

### Task 1

1. For each year, return the 3 most popular girl names and 3 most popular boy names

| Year ^ | Gender | Name        | num_babies | popularity |
|--------|--------|-------------|------------|------------|
| 1980   | F      | Jennifer    | 58379      | 1          |
| 1980   | F      | Amanda      | 35819      | 2          |
| 1980   | F      | Jessica     | 33923      | 3          |
| 1980   | M      | Michael     | 68680      | 1          |
| 1980   | M      | Christopher | 49091      | 2          |
| 1980   | M      | Jason       | 48170      | 3          |

| Year v | Gender | Name     | num_babies | popularity |
|--------|--------|----------|------------|------------|
| 2009   | F      | Isabella | 22289      | 1          |
| 2009   | F      | Emma     | 17889      | 2          |
| 2009   | F      | Olivia   | 17429      | 3          |
| 2009   | M      | Jacob    | 21162      | 1          |
| 2009   | M      | Ethan    | 19839      | 2          |
| 2009   | M      | Michael  | 18927      | 3          |

I separated the rows and ranked them in popularity by year and gender. I was able to find the 3 most popular boy and girl names for each year. I just posted the earliest and most recent years above.

## Task 2

1. For each decade, return the 3 most popular girl names and 3 most popular boy names by decade

| decade        | Gender | Name        | num_babies | popularity |
|---------------|--------|-------------|------------|------------|
| Eighties      | F      | Jessica     | 469452     | 1          |
| Eighties      | F      | Jennifer    | 440845     | 2          |
| Eighties      | F      | Amanda      | 369705     | 3          |
| Eighties      | M      | Michael     | 663645     | 1          |
| Eighties      | M      | Christopher | 554838     | 2          |
| Eighties      | M      | Matthew     | 458918     | 3          |
| Nineties      | F      | Jessica     | 303079     | 1          |
| Nineties      | F      | Ashley      | 301798     | 2          |
| Nineties      | F      | Emily       | 237222     | 3          |
| Nineties      | M      | Michael     | 462302     | 1          |
| Nineties      | M      | Christopher | 360196     | 2          |
| Nineties      | M      | Matthew     | 351596     | 3          |
| Two_Thousands | F      | Emily       | 223640     | 1          |
| Two_Thousands | F      | Madison     | 193112     | 2          |
| Two_Thousands | F      | Emma        | 181195     | 3          |
| Two_Thousands | M      | Jacob       | 273746     | 1          |
| Two_Thousands | M      | Michael     | 250471     | 2          |
| Two_Thousands | M      | Joshua      | 231851     | 3          |

I used a case when for categorizing the decades. I separated the rows and ranked them in popularity by decade and gender. I was able to find the 3 most popular boy and girl names for each decade.

### Objection 3: Compare popularity across regions

#### Task 1

1. Return the number of babies born in each of the 6 regions.

| clean_region | num_babies |
|--------------|------------|
| South        | 34219920   |
| Midwest      | 19245081   |
| Pacific      | 17540716   |
| Mid_Atlantic | 13742667   |
| Mountain     | 6282217    |
| New_England  | 4269213    |
| Midest       | 3431049    |

#### Task 2

1. Return the 3 most popular girl names and 3 most popular boy names within each region.

| clean_region | Gender | Name        | popularity |
|--------------|--------|-------------|------------|
| Mid_Atlantic | F      | Jessica     | 1          |
| Mid_Atlantic | F      | Ashley      | 2          |
| Mid_Atlantic | F      | Jennifer    | 3          |
| Mid_Atlantic | M      | Michael     | 1          |
| Mid_Atlantic | M      | Matthew     | 2          |
| Mid_Atlantic | M      | Christopher | 3          |
| Midwest      | F      | Jessica     | 1          |
| Midwest      | F      | Ashley      | 2          |
| Midwest      | F      | Sarah       | 3          |
| Midwest      | M      | Michael     | 1          |
| Midwest      | M      | Matthew     | 2          |
| Midwest      | M      | Joshua      | 3          |
| Mountain     | F      | Jessica     | 1          |
| Mountain     | F      | Ashley      | 2          |
| Mountain     | F      | Sarah       | 3          |
| Mountain     | M      | Michael     | 1          |
| Mountain     | M      | Joshua      | 2          |
| Mountain     | M      | Christopher | 3          |

| clean_region ▾ | Gender | Name        | popularity |
|----------------|--------|-------------|------------|
| South          | F      | Ashley      | 1          |
| South          | F      | Jessica     | 2          |
| South          | F      | Jennifer    | 3          |
| South          | M      | Christopher | 1          |
| South          | M      | Michael     | 2          |
| South          | M      | Joshua      | 3          |
| Pacific        | F      | Jessica     | 1          |
| Pacific        | F      | Jennifer    | 2          |
| Pacific        | F      | Ashley      | 3          |
| Pacific        | M      | Michael     | 1          |
| Pacific        | M      | Christopher | 2          |
| Pacific        | M      | Daniel      | 3          |
| New_England    | F      | Jessica     | 1          |
| New_England    | F      | Sarah       | 2          |
| New_England    | F      | Emily       | 3          |
| New_England    | M      | Michael     | 1          |
| New_England    | M      | Matthew     | 2          |
| New_England    | M      | Christopher | 3          |

## Objective 4: Explore unique names in the dataset

### Task 1

1. Find the 10 most popular androgynous names (names given to both females and males)

| Name        | num_genders | num_babies |
|-------------|-------------|------------|
| Michael     | 2           | 1382856    |
| Christopher | 2           | 1122213    |
| Matthew     | 2           | 1034494    |
| Joshua      | 2           | 960170     |
| Jessica     | 2           | 865046     |
| Daniel      | 2           | 824208     |
| David       | 2           | 819479     |
| Ashley      | 2           | 792865     |
| James       | 2           | 766789     |
| Andrew      | 2           | 761824     |

## **Task 2**

Find the length of the shortest and longest names and identify the most popular short names (those with the fewest characters) and long names (those with the most characters)

| Name            | num_babies |
|-----------------|------------|
| Ty              | 29205      |
| Franciscojavier | 52         |

The name with the few characters that was most popular was Ty with 2 characters and the name with the most characters that was most popular was Franciscojavier with 15 characters.

## **Task 3**

1. Find the state with the highest percent of babies named "Chris"



| State | pct_chris | ▼ |
|-------|-----------|---|
| LA    | 0.0335    |   |
| HI    | 0.0301    |   |
| NY    | 0.0296    |   |
| TX    | 0.0280    |   |
| CA    | 0.0268    |   |

Louisiana had the highest percentage.