

# PROJECT ASSIGNMENT 4

TITLE: DATA ANALYTICS RESEARCH PROJECT

DATASET: Popular Baby Names

NAME: Samyuktha Harshini Padmanabhuni

GNUM: G 01356642

MAJOR: MS in Information Systems

SEMESTER: Fall 2023

COURSE: INFS 580

## Table of Contents

ABSTRACT.....	3
INTRODUCTION.....	3
RESEARCH QUESTIONS:.....	4
LITERATURE REVIEW .....	4
METHODOLOGY .....	5
ABOUT THE DATASET:.....	5
DATA COLLECTION:.....	6
DATA CLEANING AND PRE-PROCESSING: .....	6
UNIVARIATE AND MULTI-VARIATE ANALYSIS:.....	8
ANALYSIS IN SQL.....	17
RESULTS AND DISCUSSIONS:.....	20
What are the most popular baby names in the dataset, and how have their rankings changed over the years?.....	20
Are there gender-related naming trends, and do certain names exhibit a shift in gender association over time? .....	23
Can we identify regional variations in baby names and naming preferences?.....	29
LIMITATIONS AND FUTURE RESEARCH:.....	32
CONCLUSIONS:.....	32
REFERENCES.....	34

# POPULAR BABY NAMES

George Mason University  
INFS 580

Samyuktha Harshini Padmanabhuni  
[spadman@gmu.edu](mailto:spadman@gmu.edu)

## ABSTRACT

This research delves into the exploration of a dataset encompassing popular baby names categorized by gender, ethnicity, and rankings over several years. Employing SQL for robust data manipulation, along with Python and R for advanced analytics and visualization, this analysis seeks to reveal temporal fluctuations in name popularity, potential shifts in gender associations with specific names, and regional diversities in naming practices. The significance of this research extends to its applicability in various domains. Understanding the dynamics of name popularity across time provides valuable cultural and societal insights shaped by historical events, media influences, and evolving gender dynamics.

**Keywords:** Gender-Related Naming Trends, Rankings, SQL, Python, R, Ethnicity, Popular.

## INTRODUCTION

New York, often referred to as a melting pot, boasts an incredibly diverse population attributed to its rich history of immigration and cultural vibrancy. The city serves as a global hub, attracting people from various ethnicities, backgrounds, and cultures worldwide. Names play a crucial role in shaping personal identity. Researching names can shed light on the psychological impacts of names on individuals, including self-perception, social interactions, and identity formation.

This dataset helps us in finding comprehensive insights into baby names encompassing popularity, gender associations, and historical trends, resonates deeply with any family. This dataset offers an invaluable opportunity to delve into the intricate realm of naming trends and their cultural significance. It encapsulates a rich tapestry of data featuring historical naming patterns, shifts in popularity, and associations with gender and ethnicity. This dataset acts as a window into understanding how names evolve over time, reflecting societal changes, cultural influences, and generational transitions.

The research endeavors to address pivotal questions within this dataset, aiming to unravel key insights. These inquiries revolve around understanding the temporal changes in the rankings of popular names, discerning gender-related naming trends, and identifying potential regional variations in naming preferences.

Driven by a genuine curiosity for the intricate world of naming conventions, this research embarks on a journey to unravel the nuances of popular baby names. The study seeks not only to analyze the dataset comprehensively but also to contribute meaningful insights into the societal, cultural, and temporal dimensions influencing naming practices within the diverse landscape of New York.

## Popular Baby Names

### RESEARCH QUESTIONS:

The below questions are answered here, in this report:

- What are the most popular baby names in the dataset, and how have their rankings changed over the years?
- Are there gender-related naming trends, and do certain names exhibit a shift in gender association over time?
- Can we identify regional variations in baby names and naming preferences?

Studying the dataset on popular baby names offers multifaceted benefits. It provides invaluable cultural and societal insights by revealing influences behind naming trends, including the impact of media, historical events, and shifting gender roles. Analyzing historical data aids in predicting future naming preferences, benefiting businesses in the baby products and services sector. Moreover, the dataset allows for geographic analysis, unveiling regional variations in naming trends, valuable for tailored marketing strategies. Exploring shifts in name associations with gender fosters discussions on identity and gender fluidity. Additionally, examining naming trends over time enriches our understanding of historical and cultural contexts, offering insights vital for sociological and anthropological research.

### LITERATURE REVIEW

#### 5.5 Million Rows of Baby Name Data?

The report is using the Social Security Administration (SSA) dataset. It looks at popular baby names in the US over time and across different places like states and territories. I noticed they use a similar dataset as mine, the "Popular Baby Names". The report discusses the use of the Social Security Administration baby names dataset, which contains information about baby names used to apply for social security cards between 1910 and 2013. It mentions that the dataset contains over 5.5 million rows of data and provides insights into popular baby names over the years. It also talks about comparing the popularity of different names and variations. The report mentions that by clicking on tree maps, you can filter the map to see which names are more popular in specific regions. It highlights variations in naming preferences based on geographic locations, such as the prominence of the name "Maria" in places with a higher Hispanic population and the regional popularity of names like "Florian" in areas with a German population. This addresses the question of regional variations in baby names. [1]

#### What Baby Names Tells Us (About Ethnic and Gender Trends)

The second report discusses the issue of gender-related naming trends and the idea of giving children unisex names. It mentions that New Yorkers prefer names that differentiate between boys and girls but cross-cultural boundaries, which relates to the concept of gender-related naming trends. While the report doesn't explicitly discuss regional variations in baby names, it talks about the multi-cultural environment of New York City and how names are becoming less ethnically specific, indicating trends in naming preferences influenced by cultural diversity. Overall, the report's content is relevant to your research questions about baby names and naming trends, particularly gender-related naming trends and the influence of cultural diversity in naming preferences. [2]

## Popular Baby Names

### What's the secrets behind most popular baby's name in New York ?

The third report mentions the importance of knowing the most popular baby names and their trends over the years. It discusses the idea of visualizing popular names to make the process of choosing a name easier. The report provides an example of the name "Angel" and how it is associated with both male and female genders and mentions its cultural and ethnic variations. This relates to the question about gender-related naming trends and shifts in gender association. It does highlight the significance of choosing common and proper names, which could be related to regional naming preferences to some extent. I used this report to give more background to my data and understand why people choose the names they do. [3]

## METHODOLOGY

### ABOUT THE DATASET:

The dataset titled "Popular Baby Names," sourced from the Department of Health and Mental Hygiene (DOHMH), provides comprehensive insights into naming trends based on civil birth registration data. It includes details such as the year of birth, gender, ethnicity, and the first name of the child. Each record within the dataset represents the ranking of a baby name based on its frequency of occurrence. The data offer a clear representation of the popularity of each name, as these ranks may fluctuate from year to year. The dataset allows for analysis of naming patterns over time, considering variations in gender, ethnicity, and the frequency of names. The dataset has 57582 rows and 6 columns. [4]

Year of Birth	int64
Gender	object
Ethnicity	object
Child's First Name	object
Count	int64
Rank	int64
<b>dtype:</b>	<b>object</b>

Figure 1 Data Types of the Dataset

The dataset has Integer and Object as data types. Also here, Ethnicity of the dataset refers to the child's Mother's Ethnicity. The number of babies with that name is coined as Count. Rank is termed as Frequency of baby names in descending order.

The dataset can be categorized into NOIR datatypes as follows:

Year of Birth: This is an **interval** variable because it has equal intervals between values, but no true zero point. For example, the difference between 2012 and 2011 is the same as the difference between 2017 and 2016, but 0 does not mean the absence of year.

Gender: This is a **nominal** variable because it is used to label categories without any quantitative value. The dataset has, female and male.

Ethnicity: This is also a **nominal** variable because it is used to label categories without any quantitative value. For example, Hispanic, Asian and Pacific Islanders, Black Non Hispanic, etc.

## Popular Baby Names

Child's First Name: This is another **nominal** variable, because it is used to label categories without any quantitative value. For example, Geraldine, Gia, Gianna, etc.

Count: This is a **ratio** variable because it has equal intervals between values, and a true zero point. For example, the difference between 13 and 10 is the same as the difference between 10 and 7, and 0 means the absence of count.

Rank: This is an **ordinal** variable because it is used to rank categories in a specific order, but the intervals between values are not equal. For example, the difference between rank 1 and rank 2 is not the same as the difference between rank 2 and rank 3.

### DATA COLLECTION:

My deep fascination with data related to New York has led me to explore various datasets within this category, but one dataset holds a particularly special place in my heart. This dataset on popular baby names strikes a personal chord with me due to a recent family experience. The arrival of my sister's newborn daughter spurred a quest for the perfect name, prompting extensive research across diverse sources for inspiration. I found this dataset in <https://catalog.data.gov/dataset/popular-baby-names> . [5]

### DATA CLEANING AND PRE-PROCESSING:

The raw dataset was totally random with different Ethnicities with different years. This screenshot is a preview of the initial dataset before cleaning. To clean the data, several methods are used.

The data was first searched to see if there were any null values. Then the dataset was made into a single font and style. It was noticed after summary statistics that few of the Ethnicities are repeated, and it possesses the same information, twice. After this initial cleaning, duplicate data was checked.

	Year of Birth	Gender	Ethnicity	Child's First Name	Count	Rank
0	2011	FEMALE	HISPANIC	GERALDINE	13	75
1	2011	FEMALE	HISPANIC	GIA	21	67
2	2011	FEMALE	HISPANIC	GIANNA	49	42
3	2011	FEMALE	HISPANIC	GISELLE	38	51
4	2011	FEMALE	HISPANIC	GRACE	36	53
...	...	...	...	...	...	...
57577	2014	MALE	WHITE NON HISPANIC	Yousef	18	94
57578	2014	MALE	WHITE NON HISPANIC	Youssef	24	88
57579	2014	MALE	WHITE NON HISPANIC	Yusuf	16	96
57580	2014	MALE	WHITE NON HISPANIC	Zachary	90	39
57581	2014	MALE	WHITE NON HISPANIC	Zev	49	65

Figure 2 Preview of the Dataset

The below screenshot shows the code to remove the null values if there are any present in the dataset.

```
#check for null values in the dataset
null = data.isnull()
print(null)

any_null = null.any().any()
print(any_null)
```

Figure 3 Code - Null values

## Popular Baby Names

It is found that **False** as the output, which says, there are null values that needs to be changed or deleted.

```
#capitalise every row and column in the dataframe
#data_cap = data.applymap(lambda x: x.capitalize() if isinstance(x, str) else x)

string_columns = data.select_dtypes(include='object').columns
data[string_columns] = data[string_columns].apply(lambda x: x.str.capitalize() if x.dtype == 'object' else x)

print(data)
```

Figure 4 Code - To Capitalize

	Year of Birth	Gender	Ethnicity	Child's First Name	Count	Rank
0	2011	FEMALE	HISPANIC	GERALDINE	13	75
1	2011	FEMALE	HISPANIC	GIA	21	67
2	2011	FEMALE	HISPANIC	GIANNA	49	42
3	2011	FEMALE	HISPANIC	GISELLE	38	51
4	2011	FEMALE	HISPANIC	GRACE	36	53
...	...	...	...	...	...	...
57577	2014	MALE	WHITE NON HISPANIC	Yousef	18	94
57578	2014	MALE	WHITE NON HISPANIC	Youssef	24	88
57579	2014	MALE	WHITE NON HISPANIC	Yusuf	16	96
57580	2014	MALE	WHITE NON HISPANIC	Zachary	90	39
57581	2014	MALE	WHITE NON HISPANIC	Zev	49	65
[57582 rows x 6 columns]						

Figure 5 Dataset before the cleaning

The above code and image show the dataset has no proper same font Capitalization.

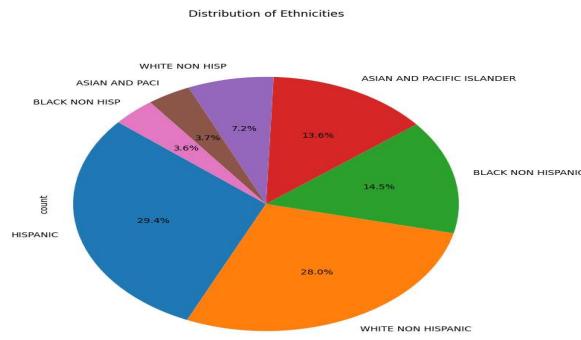


Figure 6 Distribution of Ethnicities

This pie diagram indicates irregularities in Ethnicities and so the below code is used to replace the excessive Ethnicities. Here, WHITE NON HISP and WHITE NON-HISPANIC represent the same Ethnicity.

## Popular Baby Names

```
data['Ethnicity'] = data['Ethnicity'].replace(  
    {  
        'Black non hisp' : 'Black non hispanic',  
        'White non hisp' : 'White non hispanic',  
        'Asian and paci' : 'Asian and pacific islanders',  
        'Asian and pacific islander': 'Asian and pacific islanders'  
    }  
)
```

Figure 7 Code - to replace duplicate Ethnicities.

```
#remove duplicates  
# Remove duplicate rows based on all columns  
datad = data.drop_duplicates()  
print(datad)
```

Figure 8 Code - to remove duplicates in entire dataset.

	Year of Birth	Gender	Ethnicity	Child's First Name	Count	Rank
0	2011	Female	Hispanic	Geraldine	13	75
1	2011	Female	Hispanic	Gia	21	67
2	2011	Female	Hispanic	Gianna	49	42
3	2011	Female	Hispanic	Giselle	38	51
4	2011	Female	Hispanic	Grace	36	53
...	...	...	...	...	...	...
49507	2018	Female	Black non hispanic	Nala	10	40
49508	2018	Female	Black non hispanic	Royalty	10	40
49509	2018	Female	Black non hispanic	Sariyah	10	40
49510	2018	Female	Black non hispanic	Simone	10	40
49511	2018	Female	Black non hispanic	Zahra	10	40
[18053 rows x 6 columns]						

Figure 9 After Cleaning

This is the final screenshot after cleaning. With this cleaning, initial pre-processing of data is completed.

The identification of similar ethnicities represented differently, such as "WHITE NON HISP" and "WHITE NON-HISPANIC," necessitated a consolidation step to rectify these discrepancies. The dataset underwent modifications to merge similar ethnicities, enhancing the dataset's consistency and coherence.

Visualizations, such as the pie diagram, aided in recognizing irregularities and guided the replacement process for excessive ethnicities, ensuring a streamlined dataset.

The culmination of these cleaning and pre-processing efforts resulted in a refined dataset, harmonizing ethnicities, and rectifying inconsistencies. While the initial dataset presented challenges in terms of duplicated and inconsistent entries, the systematic cleaning approach facilitated the creation of a more coherent and reliable dataset ready for subsequent analysis and exploration.

## UNIVARIATE AND MULTI-VARIATE ANALYSIS:

### Summary Stats in Python

## Popular Baby Names

```
# Assuming 'df' is the DataFrame containing the dataset
summary_stats = df.describe()

# Display summary statistics
print(summary_stats)

# Count the number of occurrences for each gender
gender_counts = df['Gender'].value_counts()

# Display the counts
print(gender_counts)
```

Figure 10 Code - summary stats py

The summary statistics generated using Python's Pandas library provided essential insights into the dataset. It was observed that the dataset spans from the year 2011 to 2019 in terms of Year of Birth, indicating the temporal range covered. Notably, numerous rows exhibited a Rank 1, signifying the prevalence of certain names as highly popular within the dataset.

	Year of Birth	Count	Rank
count	18053.000000	18053.000000	18053.000000
mean	2014.980059	33.573589	56.274193
std	2.559148	38.672649	25.503284
min	2011.000000	10.000000	1.000000
25%	2013.000000	13.000000	37.000000
50%	2015.000000	20.000000	57.000000
75%	2017.000000	35.000000	78.000000
max	2019.000000	426.000000	102.000000

Figure 11 summary stats - yob, count, rank

However, despite these findings, the summary statistics alone did not yield significantly discernible patterns or distinctive insights. The dataset comprises numerous names across varying counts, eventually categorized into different ranks. This diversity in naming occurrences resulted in limited distinct trends or noteworthy observations solely from the summary statistics.

The subsequent exploration involved visualizing the data through a box plot, aiming to present a more illustrative depiction of the distribution of names across ranks. Additionally, the analysis also involved a gender-based division, highlighting 9179 female and 8874 male occurrences within the dataset, providing a balanced representation across genders.

```
Gender
Female    9179
Male      8874
Name: count, dtype: int64
```

The gender-based division illustrated a relatively balanced representation, contributing to a more comprehensive analysis of the dataset beyond basic statistics.

## Summary Stats in R

## Popular Baby Names

```
'data.frame': 18053 obs. of 6 variables:  
$ Year.of.Birth      : int 2011 2011 2011 2011 2011 2011 2011 2011 2011 2011 ...  
$ Gender            : chr "Female" "Female" "Female" "Female" ...  
$ Ethnicity         : chr "Hispanic" "Hispanic" "Hispanic" "Hispanic" ...  
$ Child.s.First.Name: chr "Geraldine" "Gia" "Gianna" "Giselle" ...  
$ Count             : int 13 21 49 38 36 26 126 14 17 17 ...  
$ Rank              : int 75 67 42 51 53 62 8 74 71 71 ...
```

Figure 12 Data description in R

```
> summary(data)  
Year.of.Birth      Gender          Ethnicity        Child.s.First.Name  
Min.   :2011    Length:18053    Length:18053    Length:18053  
1st Qu.:2013    Class  :character  Class  :character  Class  :character  
Median  :2015    Mode   :character  Mode   :character  Mode   :character  
Mean    :2015  
3rd Qu.:2017  
Max.   :2019  
Count           Rank  
Min.   : 10.00  Min.   : 1.00  
1st Qu.: 13.00  1st Qu.: 37.00  
Median  : 20.00  Median  : 57.00  
Mean    : 33.57  Mean    : 56.27  
3rd Qu.: 35.00  3rd Qu.: 78.00  
Max.   :426.00  Max.   :102.00
```

Figure 13 Data Summary in R

The summary statistics conducted in R provided crucial insights into the dataset's structure and characteristics. The data types of various columns were identified, revealing that 'Year of Birth,' 'Count,' and 'Rank' are integer types, while 'Gender,' 'Ethnicity,' and 'Child's First Name' are characterized as strings or characters.

The statistical overview unveiled significant information about naming trends within ethnic groups. Notably, the dataset contains the count distribution of different ethnicities, emphasizing the varying prevalence of names across diverse cultural backgrounds. The least used name within the dataset has a count of 10, indicating a minimum count threshold for included names.

```
> data %>%  
+   group_by(Gender) %>%  
+   summarise(count = n())  
# A tibble: 2 × 2  
  Gender     count  
  <chr>     <int>  
1 Female     9179  
2 Male       8874  
> data %>%  
+   group_by(Ethnicity) %>%  
+   summarise(count = n())  
# A tibble: 4 × 2  
  Ethnicity      count  
  <chr>        <int>  
1 Asian and pacific islanders 3153  
2 Black non hispanic      3207  
3 Hispanic          5180  
4 White non hispanic     6513
```

Figure 14 Summary Stats - Gender, Ethnicity

## Popular Baby Names

Further exploration showcased the count distribution across specific ethnicities, with 3153 occurrences attributed to Asian and Pacific Islanders, 3207 to Black non-Hispanic, 5180 to Hispanic, and 6513 to White non-Hispanic ethnicities. This distribution highlights the prevalence and representation of different ethnic groups within the dataset, providing insights into naming preferences among diverse cultural segments.

### Univariate Analysis in R

```
> # Count the occurrences of each ethnicity
> ethnicity_counts <- table(data$Ethnicity)
>
> # Create a pie chart
> pie(ethnicity_counts, labels = names(ethnicity_counts), col = rainbow(length$+ main = "Ethnicity Distribution")
> []
> # Add a legend
> legend("topright", legend = names(ethnicity_counts), fill = rainbow(length(e$>
```

Figure 15 Code - pie, Ethnicity

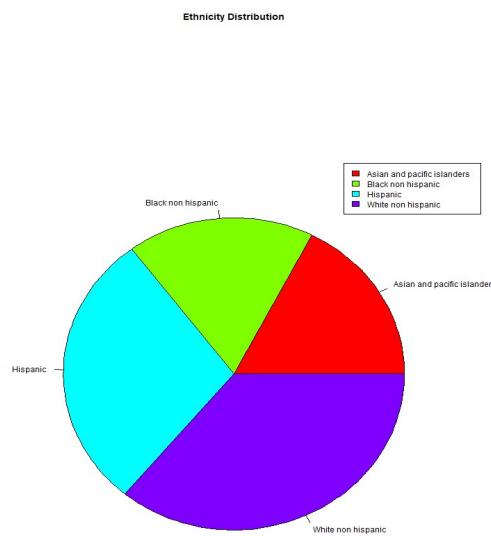


Figure 16 Ethnicity Distribution

The use of different colors (generated by the `rainbow()` function) helps distinguish between the various ethnic groups, aiding in visually comparing their proportional representation. The legend positioned at the top-right corner further clarifies the colors assigned to each ethnicity category, facilitating easy interpretation of the pie chart.

## Popular Baby Names

Overall, this pie diagram offers a clear visual depiction of the distribution of ethnicities within the dataset, highlighting the varying frequencies of different ethnic groups and emphasizing the predominant presence of White Non-Hispanic individuals, followed by Hispanic individuals, and the relatively lower count of Asian and Pacific Islanders.

```
> # Create a bar plot for Gender
> ggplot(data, aes(x = Gender)) +
+   geom_bar(fill = "skyblue", color = "black") +
+   labs(title = "Distribution of Gender", x = "Gender", y = "Count") +
+   theme_minimal()
```

Figure 17 Code - bar Plot, Gender

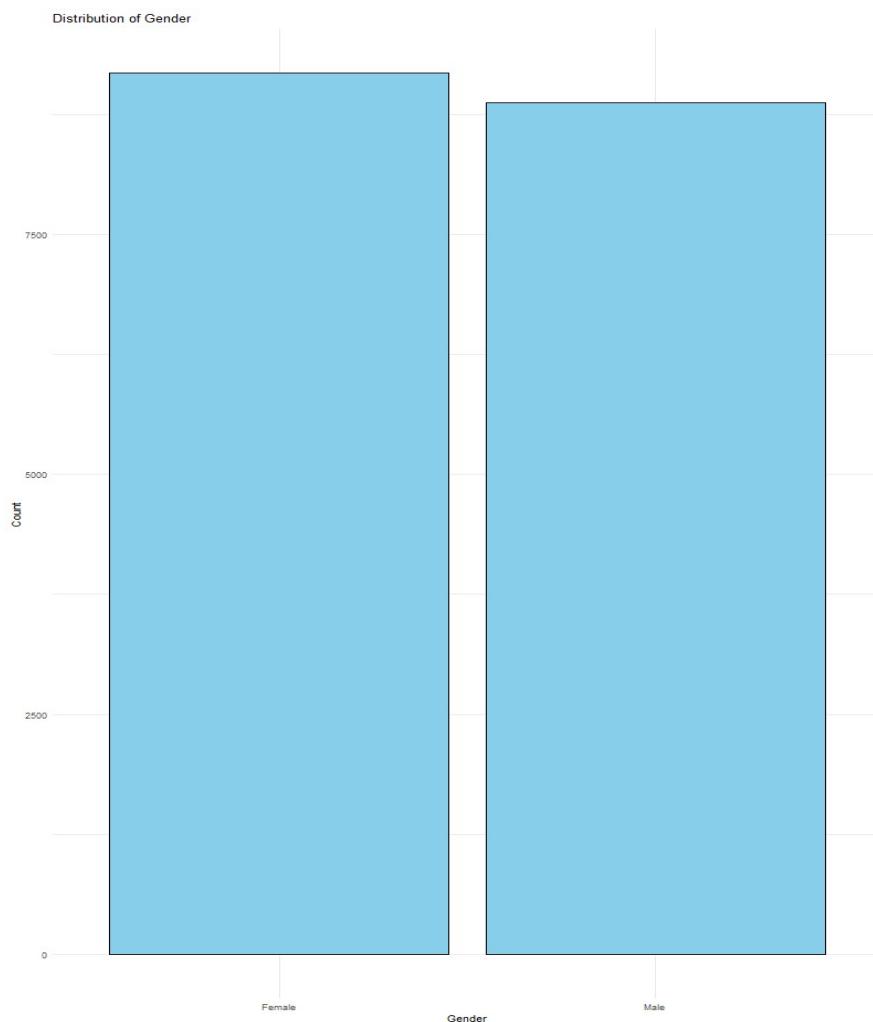


Figure 18 Distribution of Gender

The distribution of Gender is not exactly equal, Female names are more in the dataset.

## Popular Baby Names

```
# Filter only numeric columns
numeric_columns = datad.select_dtypes(include='number')
print(numeric_columns)

column = numeric_columns.columns[0]
plt.figure(figsize=(6, 4))
plt.boxplot(datad[column].dropna())
plt.title(f'Boxplot of {column}')
plt.xlabel(column)
plt.ylabel('Value')
plt.tight_layout()
plt.show()
```

Figure 19 Code - Boxplot, yob

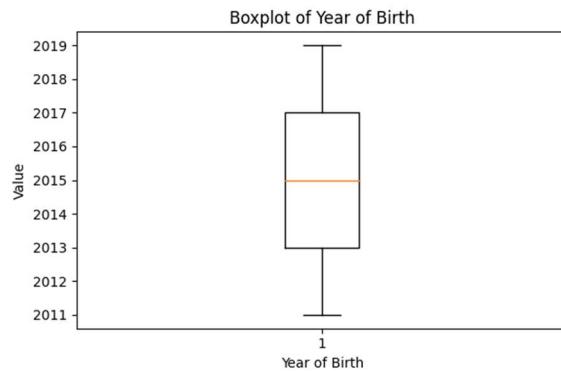


Figure 20 Year of Birth, Box plot

The box plot helps to visualize the spread, central tendency, and any potential outliers or variability in the 'Year of Birth' column, providing a clear representation of the dataset's birth year distribution. This visualization aids in understanding the distribution and variability of birth years within the dataset.

## Multivariate Analysis in Python

```
plt.figure(figsize=(12, 6))
sns.countplot(data=data, x='Year of Birth', hue='Ethnicity')
plt.title('Count Plot of Year of Birth by Ethnicity')
plt.xlabel('Year of Birth')
plt.ylabel('Count')
plt.legend(title='Ethnicity', bbox_to_anchor=(1.05, 1), loc='upper')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Figure 21 Code - Count Plot, Yob, Ethnicity

## Popular Baby Names

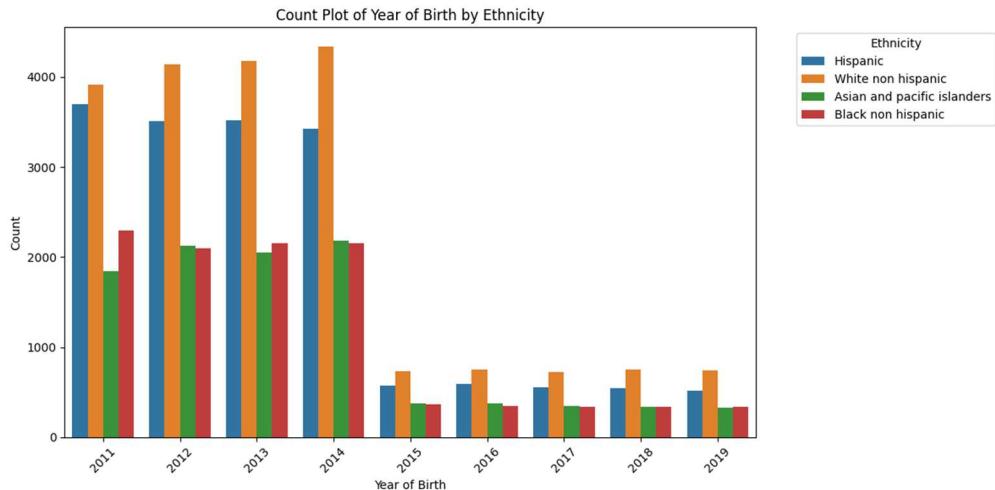


Figure 22 Count Plot of Year of Birth by Ethnicity

- The count plot allows for a visual comparison of ethnicities across multiple years. It depicts the distribution and relative counts of different ethnic groups for each year from 2011 to 2019.
- Each bar represents the count of a specific ethnicity for a particular year. Varied heights of bars indicate the frequency of each ethnicity within the dataset for different years.
- The plot's use of different colors (denoted by the legend) distinguishes between various ethnic groups, facilitating an easy comparison of their counts across the years.
- The x-axis displays the years, while the y-axis indicates the count or frequency of each ethnicity, providing an insightful view of how ethnic distribution varies across different years within the dataset.

```
# Group names alphabetically and calculate counts for each gender
datad['FirstNameGroup'] = datad["Child's First Name"].str[0] # Extract first letter of the names
grouped = datad.groupby(['FirstNameGroup', 'Gender']).size().unstack().fillna(0)

# Plotting
grouped.plot(kind='bar', stacked=True)
plt.title('Count of Names Grouped Alphabetically by Gender')
plt.xlabel('Alphabets')
plt.ylabel('Count')
plt.legend(title='Gender')
plt.show()
```

Figure 23 Code - Group plot, Gender and Names

## Popular Baby Names

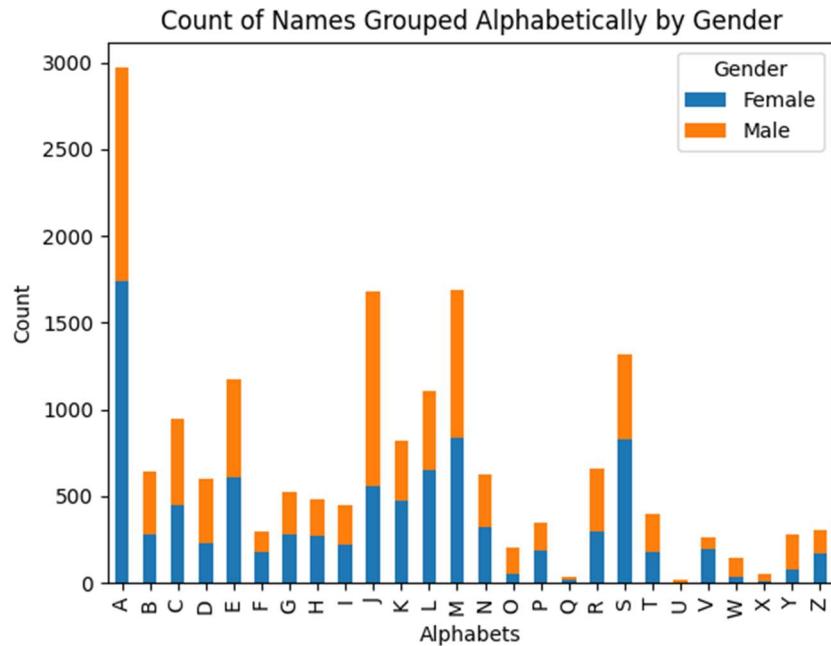


Figure 24 Group Bar Plot for different Alphabets and Genders

The plot indicates the count of names starting with different alphabets categorized by gender. The highest bars represent the alphabets ('A' and 'M') with the most substantial presence of names in the dataset. The stacked nature of the bars showcases the gender distribution within each alphabet group. The segments of different colors within each bar represent the count of names categorized by gender ('Male' and 'Female') within the respective alphabet group.

Overall, the visualization presents an overview of the distribution of names based on their initial alphabets, highlighting the count of names categorized by gender, and identifies the prevalence of specific letters at the beginning of names within the dataset.

### Multivariate Analysis in R

```
# Create a bar plot showing counts of each Ethnicity by Gender
ggplot(data, aes(x = Ethnicity, fill = Gender)) +
  geom_bar(position = "dodge") +
  labs(title = "Distribution of Gender across Ethnicities", x = "Ethnicity", y = "Count") +
  theme_minimal()
```

Figure 25 Code - Bar Plot, Ethnicity and Count

## Popular Baby Names

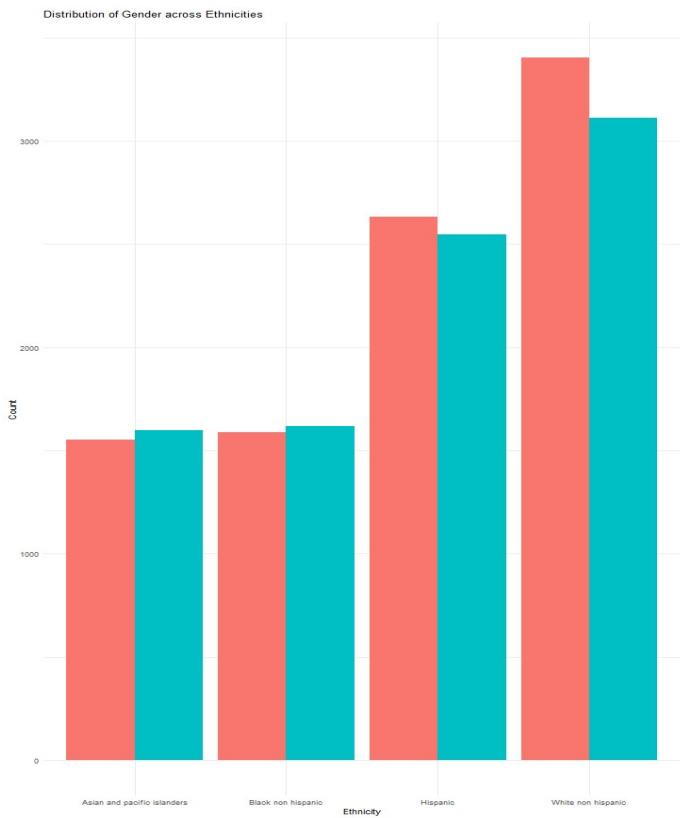


Figure 26 Bar plot - Ethnicity, Gender

This visualization effectively illustrates the gender distribution disparity across different ethnicities, emphasizing contrasting patterns of gender representation within distinct ethnic categories.

The grouped bar plot visualizes the distribution of gender ('Female' and 'Male') within various ethnicities.

In the ethnicities 'Hispanic' and 'White non-Hispanic', the count of 'Female' data appears notably higher compared to 'Male' data, indicating a larger representation of females within these ethnic groups.

Conversely, for the other two ethnicities (not explicitly mentioned in the provided information), there seems to be a higher count of 'Male' data compared to 'Female' data, implying a predominance of males within those ethnic categories.

## Popular Baby Names

### ANALYSIS IN SQL

The screenshot shows a SQL editor interface with the following details:

- Toolbar:** Includes icons for file operations, search, and refresh.
- Text Area:** Displays the following SQL code:

```
1 •   SELECT * FROM df_no_duplicates
2
3   #SELECT COUNT(*) FROM df_no_duplicates
4
5   #SELECT COUNT(*) AS Female_Count FROM df_no_duplicates
6
7
8
```
- Result Grid:** A table showing data from the query. The columns are: Year of Birth, Gender, Ethnicity, Child First Name, Count, and Ranks. The data is sorted by Count in descending order. The first few rows are: (2011, Female, Hispanic, Geraldine, 13, 75), (2011, Female, Hispanic, Gia, 21, 67), (2011, Female, Hispanic, Gianna, 49, 42), (2011, Female, Hispanic, Giselle, 38, 51), (2011, Female, Hispanic, Grace, 36, 53), (2011, Female, Hispanic, Guadalupe, 26, 62), (2011, Female, Hispanic, Hailey, 126, 8), (2011, Female, Hispanic, Haley, 14, 74), (2011, Female, Hispanic, Hannah, 17, 71), (2011, Female, Hispanic, Haylee, 17, 71), (2011, Female, Hispanic, Hayley, 13, 75), (2011, Female, Hispanic, Hazel, 10, 78), (2011, Female, Hispanic, Heaven, 15, 73), (2011, Female, Hispanic, Heidi, 15, 73), (2011, Female, Hispanic, Heidy, 16, 72), (2011, Female, Hispanic, Helen, 13, 75), (2011, Female, Hispanic, Imani, 11, 77), (2011, Female, Hispanic, Ingrid, 11, 77), (2011, Female, Hispanic, Irene, 11, 77), (2011, Female, Hispanic, Iris, 10, 78), (2011, Female, Hispanic, Isabel, 28, 60), (2011, Female, Hispanic, Isabela, 10, 78), (2011, Female, Hispanic, Isabella, 331, 1), (2011, Female, Hispanic, Isabelle, 18, 70), (2011, Female, Hispanic, Isis, 13, 75), (2011, Female, Hispanic, Itzel, 27, 61), (2011, Female, Hispanic, Izabella, 23, 65).
- Bottom Bar:** Includes buttons for Result Grid, Filter Rows, Export, and Wrap Cell Content.

Figure 27 Code - display the dataset.

SQL command to display the dataset is used.

## Popular Baby Names

```
3 •   SELECT COUNT(*) FROM df_no_duplicates  
4  
5     #SELECT COUNT(*) AS Female_Count FROM d:  
6  
7  
8
```

Result Grid    Filter Rows: _____   Export:	
COUNT(*)	
▶	18053

Figure 28 Count of the entries in the dataset

This gives the total number of entries of the dataset.

```
5 •   SELECT COUNT(*) AS Female_Count FROM df_no_duplicates WHERE Gender = 'FEMALE' group by "Year Of Birth"  
6  
7  
8
```

Result Grid    Filter Rows: _____   Export:  Wrap Cell Content:	
Female_Count	
▶	9179

Figure 29 Count of Female Entries

```
5 •   SELECT COUNT(*) AS Male_Count FROM df_no_duplicates WHERE Gender = 'MALE' group by "Year Of Birth"  
6  
7  
8
```

Result Grid    Filter Rows: _____   Export:  Wrap Cell Content:	
Male_Count	
▶	8874

Figure 30 Count of Male Entries

SQL is used to get the count of the individual gender's entries.

## Popular Baby Names

```
1 •   SELECT * FROM df_no_duplicates WHERE Ranks = 1
2
3   #SELECT COUNT(*) FROM df_no_duplicates
4
5   #SELECT COUNT(*) AS Male_Count FROM df_no_duplicates WHERE Gender =
6
7
8
```

Result Grid | Filter Rows: \_\_\_\_\_ | Export: \_\_\_\_\_ | Wrap Cell Content: \_\_\_\_\_

Year of Birth	Gender	Ethnicity	Child First Name	Count	Ranks
2011	Female	Hispanic	Isabella	331	1
2011	Female	White non hispanic	Esther	224	1
2011	Male	Asian and pacific islanders	Ethan	177	1
2011	Male	Black non hispanic	Jayden	184	1
2011	Female	Asian and pacific islanders	Sophia	119	1
2011	Female	Black non hispanic	Madison	176	1
2011	Male	White non hispanic	Michael	292	1
2011	Male	Hispanic	Jayden	426	1
2012	Female	Asian and pacific islanders	Chloe	172	1
2012	Female	Black non hispanic	Madison	159	1
2012	Female	Hispanic	Isabella	327	1
2012	Female	White non hispanic	Emma	228	1
2012	Male	Asian and pacific islanders	Ryan	197	1
2012	Male	Black non hispanic	Jayden	171	1
2012	Male	Hispanic	Jayden	364	1
2012	Male	White non hispanic	Joseph	300	1
2013	Male	Hispanic	Jayden	352	1
2017	Female	Asian and pacific islanders	Olivia	136	1
2018	Male	Hispanic	Liam	399	1
2013	Female	Asian and pacific islanders	Sophia	124	1
2013	Female	White non hispanic	Olivia	233	1
2013	Female	Black non hispanic	Madison	134	1
2014	Male	Hispanic	Liam	312	1
2013	Female	Hispanic	Isabella	326	1
2013	Male	Asian and pacific islanders	Jayden	220	1
2013	Male	White non hispanic	David	304	1
2013	Male	Black non hispanic	Ethan	146	1
2014	Female	Asian and pacific islanders	Olivia	144	1

df\_no\_duplicates 8 ×

Figure 31 Display of entries whose Rank is 1 for all years and ethnicities.

Display of Popular names in SQL, it coincides with the findings, found in Python.

The SQL commands employed various operations on the dataset named df\_no\_duplicates. The initial command SELECT \* FROM df\_no\_duplicates retrieved the entire cleaned dataset, displaying all columns

## Popular Baby Names

and rows. Utilizing COUNT (\*) function, subsequent commands calculated the total count of records present in the dataset and segregated counts based on gender ('FEMALE' and 'MALE'), grouping these counts by the 'Year of Birth'. Additionally, the command SELECT \* FROM df\_no\_duplicates WHERE Ranks = 1 extracted record representing the most popular baby names, specifically those with a rank of 1. These commands collectively facilitated data analysis by retrieving specific information such as total counts, gender-based counts per year, and identification of the most popular baby names based on their rank, enabling deeper insights into the dataset.

### RESULTS AND DISCUSSIONS:

What are the most popular baby names in the dataset, and how have their rankings changed over the years?

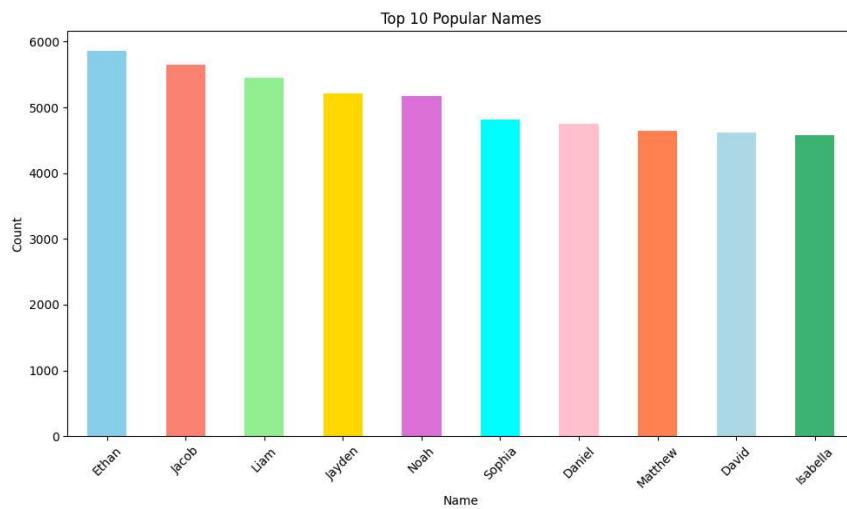


Figure 32 Popular Names

```
# Grouping data by 'Year of Birth', 'Ethnicity', 'Gender', and finding the lowest rank for each group
lowest_ranks = data.groupby(['Year of Birth', 'Ethnicity', 'Gender'])['Rank'].idxmin()

# Getting the names corresponding to the lowest ranks
most_popular_names = data.loc[lowest_ranks][['Year of Birth', 'Ethnicity', 'Gender', "Child's First Name", 'Count']]
most_popular_names
```

✓ 0.0s

Python

Year of Birth	Ethnicity	Gender	Child's First Name	Count
1085	2011	Asian and pacific islanders	Female	Sophia 119
598	2011	Asian and pacific islanders	Male	Ethan 177
1210	2011	Black non hispanic	Female	Madison 176
793	2011	Black non hispanic	Male	Jayden 184
22	2011	Hispanic	Female	Isabella 331
...	...	...	...	...
15769	2019	Black non hispanic	Male	Noah 135
14438	2019	Hispanic	Female	Isabella 231
14174	2019	Hispanic	Male	Liam 423
15386	2019	White non hispanic	Female	Chaya 209
15024	2019	White non hispanic	Male	David 252

72 rows × 5 columns

Figure 33 Few Popular Names

## Popular Baby Names

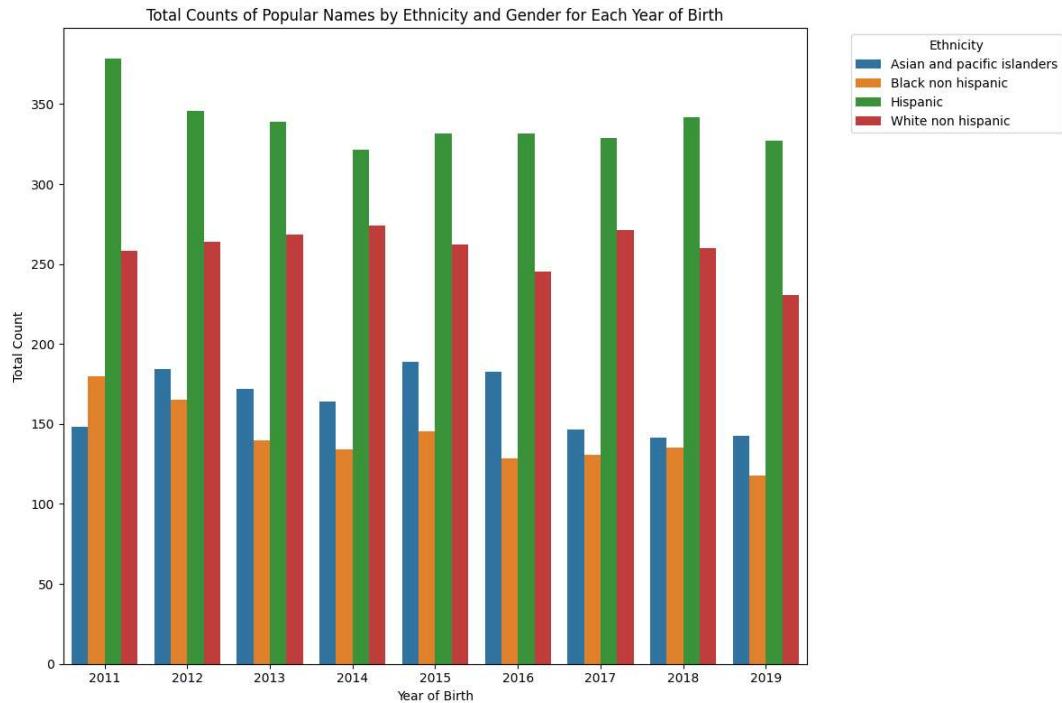


Figure 34 Count of Different Ethnicities - every year

The code performs several steps to analyze and visualize the data:

### Identifying Most Popular Names Over Years:

The code groups the data by 'Year of Birth', 'Ethnicity', and 'Gender' to find the lowest rank for each group, determining the most popular names.

A bar graph depicts the count of different ethnicities for each year, showcasing the yearly distribution of ethnicities.

```
result = data.loc[data.groupby(['Year of Birth', 'Gender'])['Count'].idxmax()]
result
✓ 0.0s
```

Year of Birth	Gender	Ethnicity	Child's First Name	Count	Rank
22	2011	Female	Hispanic	Isabella	331 1
1960	2011	Male	Hispanic	Jayden	426 1
2497	2012	Female	Hispanic	Isabella	327 1
3552	2012	Male	Hispanic	Jayden	364 1
5183	2013	Female	Hispanic	Isabella	326 1
4159	2013	Male	Hispanic	Jayden	352 1
7739	2014	Female	Hispanic	Isabella	331 1
5162	2014	Male	Hispanic	Liam	312 1
10771	2015	Female	Hispanic	Isabella	307 1
11861	2015	Male	Hispanic	Liam	356 1
12437	2016	Female	Hispanic	Isabella	276 1
11415	2016	Male	Hispanic	Liam	387 1
8713	2017	Female	White non hispanic	Esther	260 1
9447	2017	Male	Hispanic	Liam	407 1
16360	2018	Female	Hispanic	Isabella	285 1
4370	2018	Male	Hispanic	Liam	399 1
14438	2019	Female	Hispanic	Isabella	231 1
14174	2019	Male	Hispanic	Liam	423 1

Figure 35 Popular Names each Year with Rank 1

## Popular Baby Names

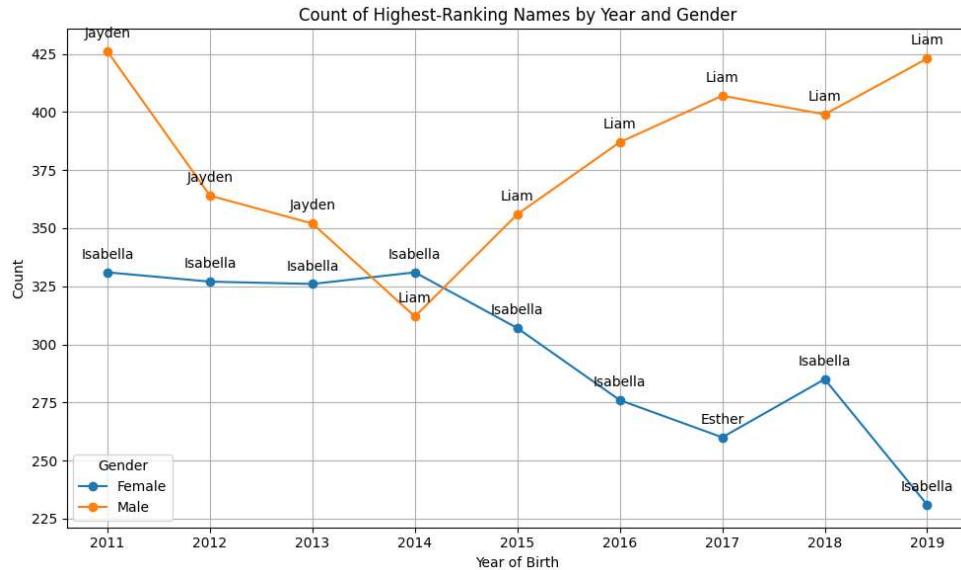


Figure 36 Representation of Rank 1 by Gender every Year

### Determining Popular Names by Gender for Each Year:

The script identifies the most popular male and female names for each year by grouping the data by 'Year of Birth' and 'Gender', extracting the names with the highest counts.

A pivot plot illustrates the highest-ranking names for males and females across various years, displaying the counts and names for each gender.

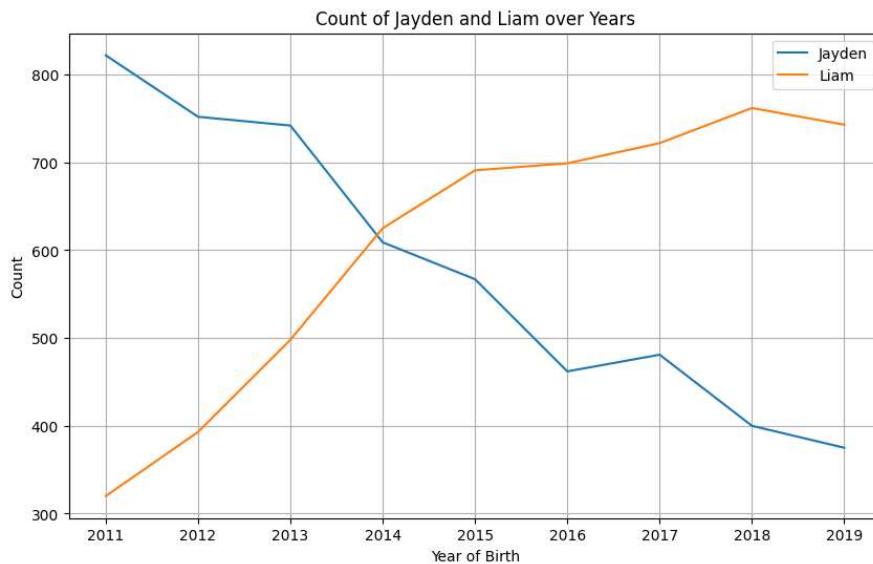


Figure 37 Exploring Jayden and Liam over years

## Popular Baby Names

### Analyzing Trends of Specific Names (Jayden and Liam):

- Focusing on specific names ('Jayden' and 'Liam'), the code filters the dataset for these names and groups the data by 'Year of Birth' to sum up their counts.
- A line plot exhibits the count trend of 'Jayden' and 'Liam' over the years, offering insights into their popularity fluctuations across different years.

The visualization demonstrates that 'Jayden' was a popular name until around 2013, followed by a surge in popularity for 'Liam' among males. Similarly, 'Isabella' emerges as a consistently popular name among females, while 'Ester' gains prominence in the year 2017. The line plot comparing 'Jayden' and 'Liam' reveals their varying counts over the years, indicating shifts in their popularity trends.

The comprehensive use of data grouping, filtering, and plotting in the code allows for a detailed exploration of popular baby names over time, facilitating an understanding of how name rankings have evolved across different years within the dataset.

To note, the research initially aimed to find popular names by considering those with rank 1 across all years. However, due to the extensive dataset and the complexity in identifying trends based solely on ranks, the focus shifted to extracting the names with the highest counts for all years. Consequently, Jayden and Liam emerged as the most popular names, allowing a clearer analysis of the changes in their popularity trends over time.

Are there gender-related naming trends, and do certain names exhibit a shift in gender association over time?

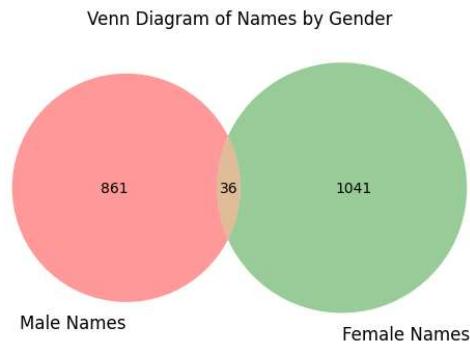


Figure 38 Venn Diagram of Names by Gender

### Data Filtering and Grouping:

- The code filters the dataset to include names that have been used both as male and female names. It creates a list of specific names known to have been used for both genders.
- After filtering the data, it groups it by 'Year of Birth', 'Gender', and 'Child's First Name' to count the occurrences of each name across genders for each year.

## Popular Baby Names

Names that appear for both genders:		
Gender	Female	Male
Child's First Name		
Alexis	19.0	9.0
Angel	12.0	12.0
Ariel	28.0	18.0
Avery	32.0	25.0
Blake	7.0	17.0
Cameron	2.0	32.0
Charlie	7.0	13.0
Dylan	10.0	36.0
Eden	14.0	8.0
Elliott	1.0	9.0
Emerson	9.0	2.0
Finley	3.0	1.0
Jamie	4.0	1.0
Jaylin	8.0	1.0
Jia	1.0	1.0
Jordan	7.0	35.0
Kai	1.0	32.0
Logan	3.0	36.0
Milan	9.0	10.0
Nana	1.0	1.0
Parker	5.0	11.0
Peyton	22.0	5.0
...		
Skyler	5.0	4.0
Taylor	23.0	2.0
Tenzin	9.0	9.0
Yael	9.0	2.0

Figure 39 Names that are used for Both Genders

```

selected_names = ['Alexis', 'Angel', 'Ariel', 'Avery', 'Blake', 'Cameron', 'Charlie', 'Dylan', 'Eden', 'Elliott', 'Emerson', 'Finley', 'Jordan', 'Kai', 'Logan', 'Milan', 'Nana', 'Parker', 'Peyton', 'Skyler', 'Taylor', 'Tenzin', 'Yael']
data['Year of Birth'] = data['Year of Birth'].astype(str)
filtered_data = data[data["Child's First Name"].isin(selected_names)]
#filtered_data = filtered_data.sort_values(by="Child's First Name")
sorted_data = filtered_data.sort_values(by=["Child's First Name", 'Ethnicity', 'gender'])
sorted_data
    ✓ 0.1s

```

Year of Birth	Gender	Ethnicity	Child's First Name	Count	Rank	
1974	2012	Female	Asian and pacific islanders	Alexis	18	37
11372	2015	Female	Asian and pacific islanders	Alexis	12	41
1112	2011	Female	Black non hispanic	Alexis	13	42
2205	2012	Female	Black non hispanic	Alexis	12	45
4540	2013	Female	Black non hispanic	Alexis	12	45
...	...	...	...	...	...	...
8911	2017	Female	White non hispanic	Yael	20	74
11665	2015	Female	White non hispanic	Yael	19	74
11937	2016	Female	White non hispanic	Yael	20	74
15551	2019	Female	White non hispanic	Yael	22	72
17617	2018	Female	White non hispanic	Yael	14	81

739 rows × 6 columns

Figure 40 Code which has all Names which are used both as Female and Male

### Analysis of Gender Association for Each Name Over Years:

- For the selected names, the code iterates through each name and year to determine if it's predominantly used as a male name, a female name, or equally for both genders.
- By comparing the total counts of the name for males and females in each year, it prints out the predominant gender association for each selected name in each year within the dataset.

## Popular Baby Names

```
In 2011, Alexis is used more as a Female name.  
In 2012, Alexis is used more as a Female name.  
In 2013, Alexis is used more as a Female name.  
In 2014, Alexis is used more as a Female name.  
In 2015, Alexis is used more as a Female name.  
In 2016, Alexis is used more as a Female name.  
In 2017, Alexis is used more as a Female name.  
In 2018, Alexis is used equally as Male and Female name.  
In 2019, Alexis is used equally as Male and Female name.  
In 2011, Angel is used equally as Male and Female name.  
In 2012, Angel is used more as a Female name.  
In 2013, Angel is used more as a Female name.  
In 2014, Angel is used more as a Male name.  
In 2015, Angel is used equally as Male and Female name.  
In 2016, Angel is used equally as Male and Female name.  
In 2017, Angel is used more as a Female name.  
In 2018, Angel is used more as a Male name.  
In 2019, Angel is used more as a Male name.  
In 2011, Ariel is used more as a Female name.  
In 2012, Ariel is used more as a Female name.  
In 2013, Ariel is used more as a Female name.  
In 2014, Ariel is used more as a Female name.  
In 2015, Ariel is used more as a Female name.  
In 2016, Ariel is used equally as Male and Female name.  
In 2017, Ariel is used more as a Female name.  
...  
In 2016, Yael is used more as a Female name.  
In 2017, Yael is used more as a Female name.  
In 2018, Yael is used equally as Male and Female name.  
In 2019, Yael is used more as a Female name.
```

Figure 41 Information which says about the usage

The analysis provides insights into how certain names' gender associations have evolved over time. For instance:

- 'Alexis' was predominantly used as a female name from 2011 to 2018, and it became equally popular for both genders in 2018 and 2019.
- 'Angel' fluctuated between being predominantly used as a female and male name across different years.
- 'Ariel' was predominantly a female name throughout the years analyzed.
- 'Charlie' showed variations but leaned more towards being a male name in several years.
- 'Dylan' consistently appeared as a male name across all years.
- 'Taylor' was predominantly a female name in all years analyzed.

This analysis sheds light on the fluctuating gender associations of certain names over the years within the dataset. Names like 'Alexis' and 'Angel' showed changes in their gender associations, while others, like 'Dylan', maintained a consistent gender association throughout the analyzed period. The results offer insights into the evolving nature of gender-related naming trends.

Every year the name is more used for Male:

## Popular Baby Names

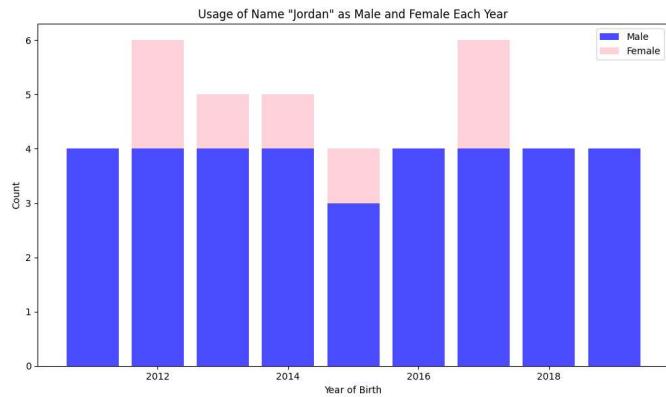


Figure 42 Jordan

Every year the name is more used for Female:

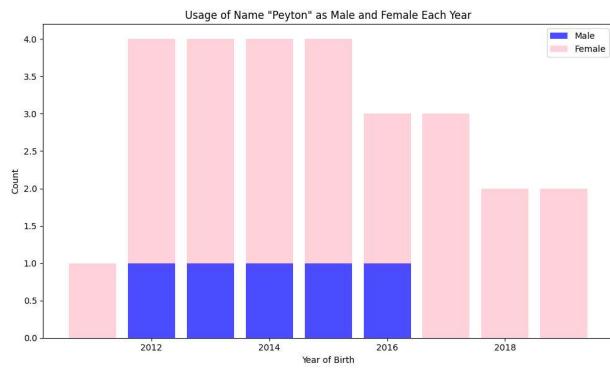


Figure 43 Peyton

Every year the name has equal usage as Male name and Female name:

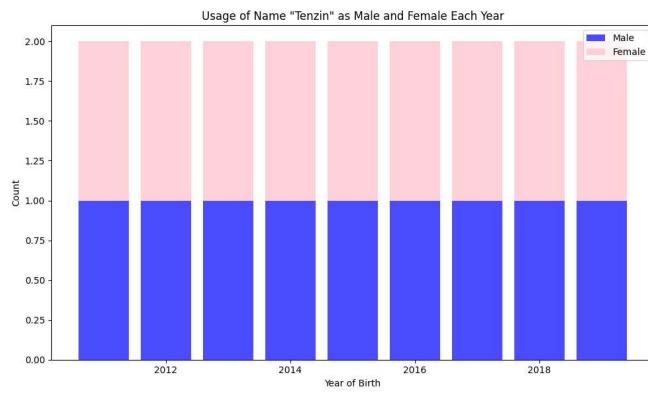


Figure 44 Tenzin

Other Combinations:

## Popular Baby Names

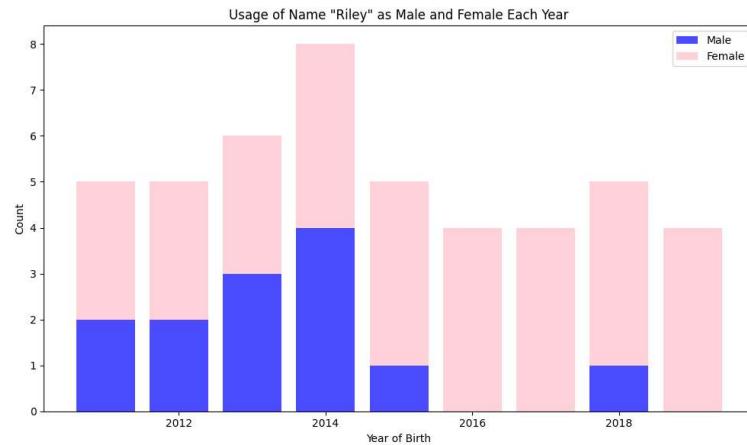


Figure 45 Riley

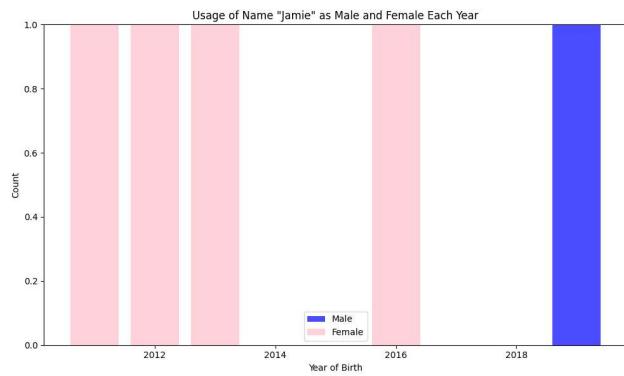


Figure 46 Jamie

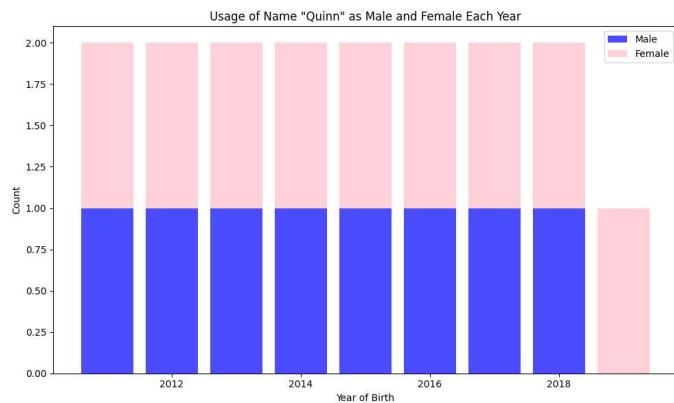


Figure 47 Quinn

## Popular Baby Names

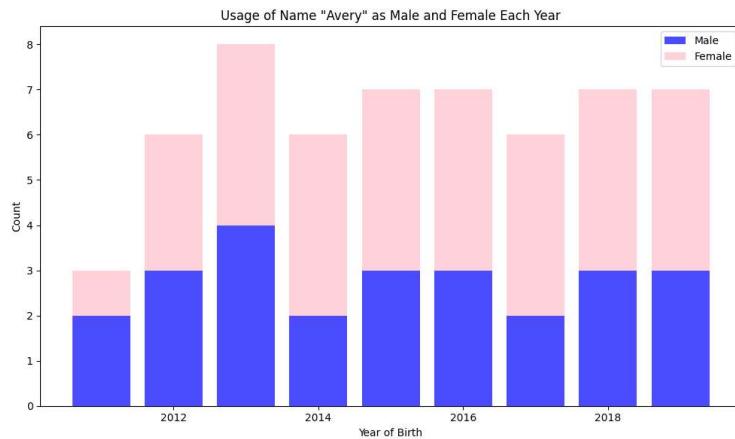


Figure 48 Avery

1. **Jordan:** From 2011 to 2019, Jordan showed a consistent trend of being predominantly used as a male name. It was more frequently associated with males across all the analyzed years.
2. **Peyton:** Throughout the years under examination, Peyton was primarily used as a female name. It consistently exhibited higher usage as a female name across the dataset.
3. **Tenzin:** Tenzin's usage demonstrated an interesting trend of being equally used as a name for both genders across the years analyzed. There wasn't a distinct prevalence as either a male or female name.
4. **Riley:** The usage of Riley appeared to shift over time. In some years, it was used equally as a name for both males and females, while in others, it leaned more towards being a female name.
5. **Jamie:** Jamie's usage showcased variation over the years. It was predominantly used as a female name in some years, while in 2019, it was more frequently associated as a male name.
6. **Quinn:** Across the analyzed years, Quinn's usage was fairly consistent, being equally employed as a name for both males and females in most years.
7. **Avery:** Avery's usage revealed a shift over the analyzed years, trending more towards being a male name initially but evolving to be predominantly associated with females in later years.

To address the research question regarding gender-related naming trends and names exhibiting shifts in gender association over time, the analysis specifically concentrated on names that were utilized as both male and female names. The exploration aimed to discern varying trends among different names and understand their gender associations across different years. Multiple visualization techniques were employed to elucidate the analysis.

It is evident that different names exhibited diverse trends over the analyzed period. Through the examination of names used for both genders, various visualizations were utilized to uncover and represent the evolving gender associations. Despite the comprehensive analysis, it became apparent that there isn't a singular, universal trend observed across all names. Each name displayed its unique pattern, depicting shifts in gender association over time. This underscores the nuanced nature of gender-related naming trends, indicating that there isn't a singular, definitive trend applicable to all names.

## Popular Baby Names

Can we identify regional variations in baby names and naming preferences?

```
sorted_data = data.sort_values(by='Count', ascending=False)
filtered_data = sorted_data[~sorted_data.duplicated(subset='Child\'s First Name')]
filtered_data.head(10)
print(filtered_data["Child's First Name"].head(10))
```

```
1960      Jayden
14174     Liam
13544     Jacob
10599     Dylan
12230     Ethan
7739      Isabella
1526      Justin
11146     Matthew
6311      David
8075      Joseph
Name: Child's First Name, dtype: object
```

Figure 49 Popular names

Year of Birth	Gender	Ethnicity	Child's First Name
622	2011	Male Asian and pacific islanders	Jayden
793	2011	Male Black non hispanic	Jayden
1771	2011	Male White non hispanic	Jayden
1960	2011	Male Hispanic	Jayden
Count	Rank		
622	173	2	
793	184	1	
1771	39	68	
1960	426	1	

Figure 50 Jayden - 2011

## Popular Baby Names

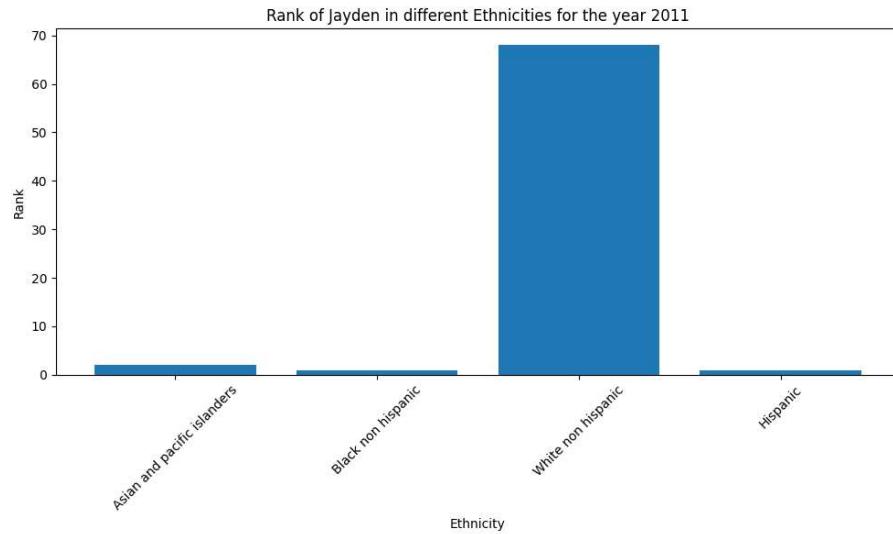


Figure 51 Ranks of Jayden in 2011

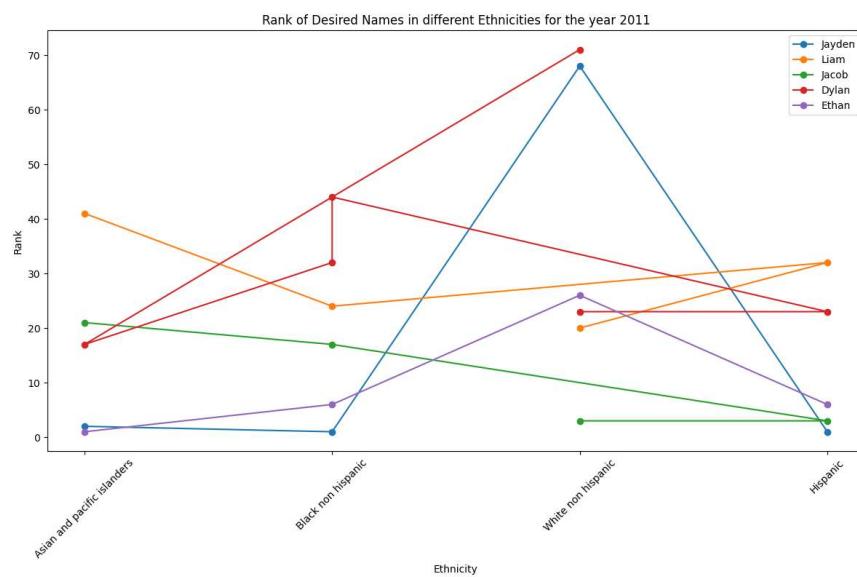


Figure 52 Exploring 5 names for the year 2011 in different Ethnicities.

## Popular Baby Names

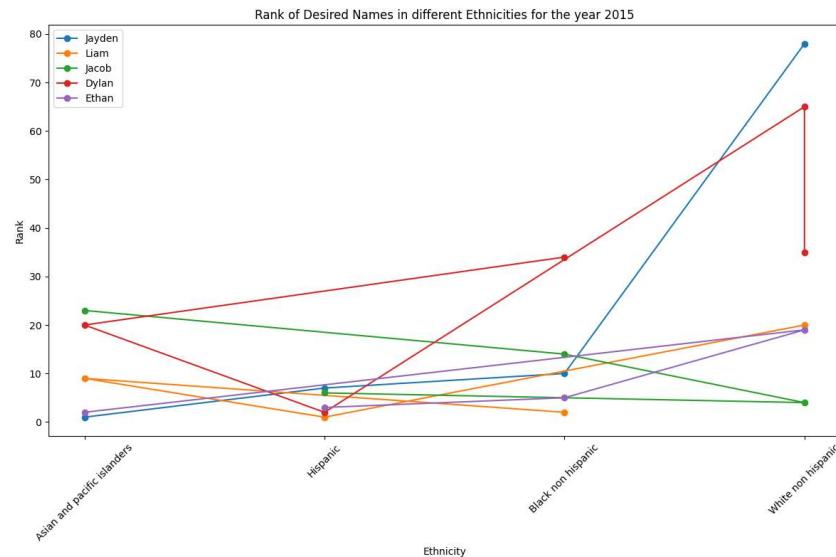


Figure 53 Exploring 5 names for the year 2015 in different Ethnicities

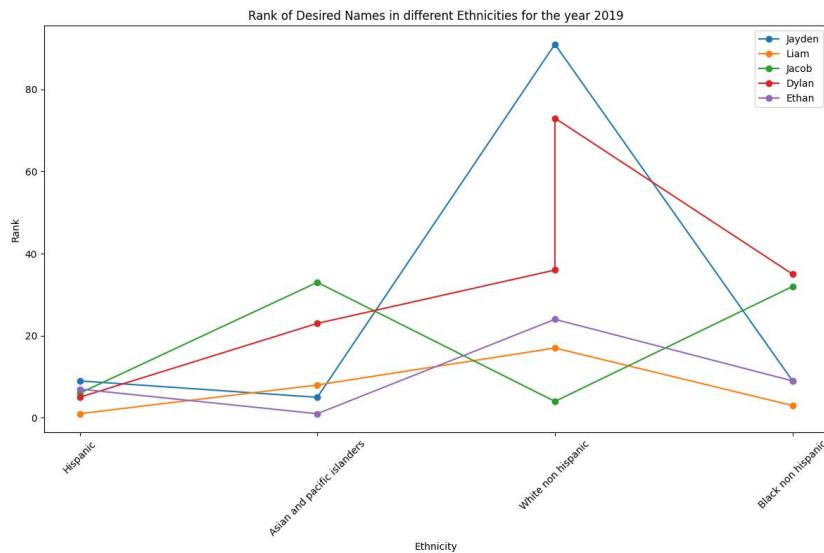


Figure 54 Exploring 5 names for the year 2019 in different Ethnicities

Based on the analysis conducted for different years (2011, 2015, and 2019) across various ethnicities and for specific names (Jayden, Liam, Jacob, Dylan, and Ethan), notable observations were made concerning regional variations in baby names and naming preferences.

In 2011, Dylan emerged as a popular name across genders and ethnicities. Jayden maintained a generally high rank across all ethnicities. Ethan had relatively lower ranks, especially in Asian and Pacific Islanders, whereas Liam had moderate rankings across ethnicities.

Moving to 2015, a more diverse scenario unfolded. Dylan was notably prominent, used for both males and females in White non-Hispanic ethnicities. Liam exhibited lower ranks across most ethnicities, while

## Popular Baby Names

Jayden had the highest rank in White non-Hispanic, surpassing others among different names and ethnicities.

By 2019, there were further shifts in naming preferences. Ethan and Liam continued to display lower ranks across ethnicities. Jayden, while not the most popular in White non-Hispanic, maintained a consistent high rank. Dylan, again, was utilized for both males and females in White non-Hispanic, showcasing its popularity in that category.

These findings illustrate considerable variations in naming preferences across different ethnicities and over time. The names experienced fluctuations in popularity, with some consistently ranking higher across ethnic groups and others displaying varying trends. Therefore, it's evident that regional variations exist in baby names and naming preferences, with certain names holding more prominence or popularity within specific ethnic communities or years.

### LIMITATIONS AND FUTURE RESEARCH:

- **Incomplete Yearly Data Equality:** The dataset exhibits gaps in providing equal data representation across all years, potentially missing or containing unequal information for certain years. This variability could hinder comprehensive year-to-year comparisons and trend analysis, affecting the accuracy of insights drawn from the dataset.
- **Ethnic Representation Limitations:** While categorizing ethnicities, the dataset overlooks certain groups, notably Asian and Pacific Islanders, where specific subgroups like Japanese individuals might be absent. This limitation in ethnic representation can skew the broader understanding of naming preferences within diverse cultural segments, leading to incomplete insights and potentially overlooking significant naming trends.
- **Regional Scope Limited to New York:** The dataset is exclusive to New York, restricting its scope to this geographic region. As a result, it lacks broader national or international perspectives, limiting the generalizability of findings to a specific locale and potentially missing out on broader societal naming trends.
- **Absence of Alphabetical Organization:** The dataset lacks information regarding the alphabetical arrangement of names, making it challenging to discern patterns or trends based on specific alphabetical sequences. This absence restricts analyses exploring naming preferences or associations within alphabetic ranges.
- **Limitation on Representation of Mother's Ethnicity (Pure or Mixed):** The dataset categorizes the ethnicity of the mothers; however, it might lack granularity regarding the representation of pure or mixed ethnic backgrounds.

### CONCLUSIONS:

The exploration of the "Popular Baby Names" dataset spanning 2011 to 2019 in New York offered profound insights into naming trends, gender associations, and regional variations. Through in-depth analysis using SQL, Python, and R, intriguing patterns emerged, shedding light on the dynamics of popular names and their changing rankings across years, the nuanced gender associations of specific names, and the identification of regional naming preferences.

The investigation primarily sought to uncover the most popular names over the years and understand their evolving trends. Initially focusing on names with rank 1 across all years, the analysis transitioned to extract names with the highest counts. Notably, Jayden and Liam emerged as the most popular names,

## Popular Baby Names

enabling a clearer examination of changes in their popularity trends. This shift in focus led to a comprehensive exploration of the popularity of specific names across different years and ethnicities.

Exploring gender-related naming trends and the fluctuating gender associations of specific names, such as 'Alexis,' 'Angel,' 'Ariel,' 'Charlie,' 'Dylan,' and 'Taylor,' provided insights into the evolving nature of naming conventions. These findings underscored the changing perceptions of gender associations in names over the years, showcasing varied trends for each name across different genders and ethnicities.

Moreover, the analysis of additional names like Jordan, Peyton, Tenzin, Riley, Jamie, Quinn, and Avery highlighted their distinct trends and gender associations, further emphasizing the diverse patterns observed in naming preferences and gender representation.

Further investigations into regional variations in baby names underscored the diversity in naming preferences across different ethnicities and years. For instance, the analysis revealed that specific names like Dylan, Jayden, Liam, Ethan, and others exhibited varying ranks across ethnic groups and years, indicating shifts in their popularity and regional preferences.

In conclusion, the comprehensive analysis of the "Popular Baby Names" dataset illuminated the intricate dynamics of naming trends, gender associations, and regional variations. The research uncovered diverse trends in naming preferences, emphasizing the unique patterns exhibited by different names across genders, ethnicities, and years. These findings underline the multifaceted nature of naming conventions and the evolving landscape of baby names, offering valuable insights into societal shifts and preferences in New York over the analyzed period.

## REFERENCES

- [1] J. Loree, "5.5 Million Rows of Baby Name Data," 2015. [Online]. Available: <https://www.tableau.com/blog/55-million-rows-baby-name>. [Accessed 31 10 2023].
- [2] "About Ethnic and Gender Trends WHAT BABY NAMES TELL US WHAT BABY NAMES TELL US," [Online]. Available: <https://cdn.sisense.com/wp-content/uploads/What-Baby-Names-Tell-Us-About-Ethnic-and-Gender-Trends.pdf>. [Accessed 31 10 2023].
- [3] S. Song, "What's the secrets behind most popular babys' name In new york ? , " 7 3 2019. [Online]. Available: <https://studentwork.prattsi.org/infovis/visualization/whats-the-secrets-behind-most-popular-babys-name-in-new-york%ef%bc%9f/>. [Accessed 31 10 2023].
- [4] N. O. Data. [Online]. Available: <https://data.cityofnewyork.us/api/views/25th-nujf/rows.json?accessType=DOWNLOAD>. [Accessed 28 11 2023].
- [5] N. O. Data. [Online]. Available: <https://catalog.data.gov/dataset/popular-baby-names>. [Accessed 25 11 2023].