Analyzing the Influence of Netflix Streaming Titles on Australian Baby Naming Trends*

Himavanth Reddy Kalakota
Master of Data Science
University of Adelaide
Adelaide, Australia
a1953735@adelaide.edu.au

***Abstract*—The fact that digital media has a significant influence on cultural behavior is now more pronounced than ever during the streaming era. This study explores whether the names of popular streaming series and movies have begun to influence baby names in Australia. Through cross-reference of the two separate open source datasets, one from a collection of Australian baby name records (2014-2024) and the other from Netflix's global catalog of TV and movie titles, the project hopes to find correlations in the appearance of a streaming title with the rise of babies receiving the same name. The amount of data is massive, dealing with both genders, years, and platforms, so it is a kind of big data task. The initial analysis examines exact name matching between Netflix titles and baby names and comparative trends before and after the year of release of the title. I describe a few limitations (e.g., ambiguous matches and partial name overlap), which result in a modification of the research question and a backup plan regarding characters' names. The initial analysis plan is set, and the data is analyzed using Python for the next work. This project highlights how pop culture may influence real-world decisions, particularly in parental name choices.**

I. Introduction

Nowadays, streaming platforms like Netflix, Amazon Prime Video, and Disney+ draw people towards popular culture. These platforms influence the entertainment we select, as well as how we act as a community—everything from fashion to language and naming. One particularly under-researched force is the effect of trendy streaming offerings, especially words that can function as names, on baby names.

The choice of your baby's name is both a cultural and emotional statement. It reflects your personal preference, to be sure, but then again, it is something from your family, and now with our media culture, it is something visited upon a child. There are anecdotal and numerical signs that parents might use names for their newborns that appear in popular media [1] [5]. Using the accessible data files from the open data produced by the South Australian government on baby names over time, we are able to examine whether similar associations can be supported in the context of Australia from streaming titles.

This is true for baby name data in South Australia [2] [6] as well as metadata about Netflix titles containing type, year of release, and availability [3], for instance. These data clusters are more than a decade old, and they comprise structured data in a set of files and tables in diverse formats and systems, in conformity with the notion of Big Data as described in the literature [4].

_id	First Name	Amount	Position
1	OLIVER	131	1
2	HENRY	125	2
3	NOAH	105	3
⋮	⋮	⋮	⋮
100	FELIX	19	=98

TABLE I
EXAMPLE TABLE SHOWING ALL FIELDS FROM THE BABY'S NAME DATASET [2]

The current study asks the following research question: "Do the names of popular streaming titles (TV shows and movies) influence baby name trends in South Australia?"

By comparing the names of streaming titles against baby names from 2014 to 2024, this project aims to find the themes of where culture's impact lives. It additionally features a redundancy analysis using the names of characters if influence via the title is too low.

II. DATA SOURCES

This study uses two open-source datasets. One consists of baby name records from South Australia, and the other consists of metadata on Netflix titles. These datasets were selected due to their complementary domains and potential for time-based correlation, supporting the central research question.

A. Baby's Name Data (South Australia)

The baby's name dataset was sourced from the South Australian Government's open data portal [2]. It consists of multiple comma-separated value (CSV) files, each representing one gender (male or female) for a specific year between 2014 and 2024.

Each file contains the top 100 baby names for that group, along with their frequency of occurrence.

Each entry includes the following fields as shown in Table I:

- id: It's just the numbering of the data [2]
- First Name: It is the first name of the people [2]
- Amount: It is the frequency of the first name [2]
- Position: It is the rank of the First Name [2]
- Year (inferred from the filename) [2]

- Gender (inferred from the folder structure) [2]

Each of the datasets comprises 3000 name records, enabling analysis of naming trends within South Australia.

B. Netflix Titles Dataset

The second dataset was sourced from Kaggle's public dataset repository [3]. It contains metadata on over 8,000 titles available on the Netflix platform, including both movies and TV shows.

The relevant columns for this study from the Table II are:

- show id: Unique ID for every Movie/TV Show Type(TV Show or Movie) [3]
- type: Identifier - A Movie or TV Show Date Added [3]
- director: Director of the Movie [3]
- cast: Actors involved in the movie/show [3]
- country: Country where the movie/show was produced [3]
- date added: Date it was added to Netflix [3]
- release year: Actual Release year of the movie/show [3]
- rating: TV Rating of the movie/show [3]
- duration: Total Duration - in minutes or number of seasons [3]

The dataset is provided as a single CSV file. Preliminary filtering is required to remove null values, standardize title formatting, and ensure distinct entries for accurate name matching.

C. Big Data Criteria (4Vs)

This project satisfies the commonly accepted 4Vs definition of Big Data:

- Volume: The datasets include thousands of entries covering over a decade of records.
- Variety: The project integrates two different data types: structured government birth data and semi-structured entertainment metadata.
- Velocity: Baby name data is released annually, while Netflix content metadata includes the time of platform addition, supporting temporal comparisons.
- Veracity: Both datasets originate from trusted sources, a state government repository and a well-documented, widely used public dataset.

The integration of these sources enables a robust foundation for identifying patterns in baby naming influenced by streaming media content.

III. INITIAL DATA PROCESSING AND ASSESSMENT

This section is the blueprint of the early data processing steps and provides an assessment of the data's adequacy in answering the research question.

A. Processing the Baby's Name Data

The baby's name data consisted of CSV files separated by gender and year. A total of 22 files (11 for each gender from 2014 to 2024) will be loaded using Python's pandas library. A new column, Year, can be derived from the filename, and a Gender column can be assigned based on the folder path. The relevant columns (First Name, Amount, Year, Gender) were retained, and all files were concatenated into a single DataFrame for analysis. All names were converted to lowercase to match Netflix titles. The consolidated dataset contains almost 2,000 unique name-year-gender entries.

B. Processing the Netflix Titles Dataset

The Netflix dataset was provided as a single CSV file and required filtering for records with null or empty titles. Only records with valid Title, Release Year, and Type (TV Show or Movie) fields were retained. All titles were lowercased to allow for comparison against baby name entries. A primary inspection showed that many titles were phrases or non-name words (e.g., The Social Dilemma, Money Heist), which may not be used as baby names. Therefore, only single-word titles with name-like structures were considered as potential matches. This subset was isolated for further comparison.

C. Data Adequacy and Deficiencies

Initial inspection of the two datasets revealed the following challenges:

- Low Match Rate: A small proportion of Netflix titles are also valid baby names. For example, titles like Wednesday, Moana, and Enola are name-like, but many others are not.
- Ambiguity: Some matches may be coincidental. For example, "Grace" may spike in popularity for reasons unrelated to Grace and Frankie.
- Name Structure: Multi-word titles (e.g., Anne with an E) cannot be matched directly to baby names without parsing.
- Lack of Popularity Metrics: The Netflix dataset does not include viewership or popularity, but only presence on the platform.

D. Response and Refinement Plan

To improve the validity of the analysis, the following refinements are proposed:

- Focus the analysis on a filtered list of exact-match, single-word Netflix titles.
- Incorporate time-based filtering to analyze baby name spikes following the release year of matching Netflix titles.
- Include a redundancy check: If few name-title matches exist, consider a backup research question using fictional character names instead of titles.

IV. REFINED QUESTION AND PLAN

Following the initial inspection of both datasets, the original research question has been refined to better align with the structure and limitations of the data.

show_id	Type	Title	Director	Cast	Country	Date Added	Release Year	Rating	Duration	Genre	Description
s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	–	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmmaker...
s2	TV Show	Blood & Water	–	Ama Qamata, Khosi Ngema, Gail Mabalane, ...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, ...	After crossing paths at a party, a Cape Town tee...
s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, ...	–	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, ...	To protect his family from a powerful drug lord...
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a tyco...

TABLE II
EXAMPLE TABLE SHOWING ALL FIELDS FROM THE NETFLIX TITLES DATASET [3]

A. Refined Research Question

The initial broad question was: ***"Do the names of popular streaming titles (TV shows and movies) influence baby name trends in South Australia?"***

After assessing the datasets, this has been refined to: ***"Did the release of a Netflix streaming title with a single-word name cause a noticeable change in the frequency of babies receiving that name in South Australia between 2014 and 2024?"***

This refined question narrows the scope to titles that are:

- Single word
- Likely to be used as first names
- Have clear release years for temporal comparison

B. Next Steps in Data Analysis

The project will proceed with the following steps:

1) Title-Name Matching: Determine Netflix titles that consist of single words and compare them with baby names from the SA dataset.
2) Temporal Comparison: Plot the Graph of the occurrence of each identified name before and following the release year to examine any significant increases or rising trends.
3) Trend Significance: Evaluate whether the rise (if present) is statistically meaningful or if it could simply result from random fluctuations.
4) Machine Learning Methods: Logistic regression or classification models may be applied to predict whether a given title-name match is likely to cause a naming spike (based on year, genre, etc.).
5) Clustering techniques (e.g., K-Means) could help group names with similar trend behavior.
6) Visualization: Use Python's matplotlib and seaborn to plot name trends over time and highlight release years for context.
7) Interpretation: Evaluate findings for cultural relevance, and assess whether observed changes are likely influenced by the media exposure of the streaming title.

We use the Python libraries like pandas, matplotlib, scikit-learn, and seaborn for the evaluation.

C. Backup Question and Data Source

If the number of valid name-title matches is too low to support a conclusive analysis, an alternate direction is proposed: ***"Do fictional character names from popular Netflix series and movies influence baby name trends in South Australia?"***

In this case, the dataset will be extended to include main character names from high-profile Netflix titles (e.g., "Eleven" from Stranger Things, "Nairobi" from Money Heist). Character names can be manually extracted from IMDb or Wikipedia. This backup strategy ensures the project remains viable even if exact title-name overlaps are sparse.

ACKNOWLEDGMENT

I acknowledge that this work is my own and has been completed in accordance with the academic integrity policies of the University of Adelaide.

REFERENCES

[1] Behind the Name, 2023. Name Popularity Trends. [online] Available at: https://www.behindthename.com/top/

[2] South Australian Government, 2024. Popular Baby Names - South Australia. [online] Available at: https://data.sa.gov.au/data/dataset/popular-baby-names

[3] Shivam Bansal, 2020. Netflix Movies and TV Shows. Kaggle. [online] Available at: https://www.kaggle.com/datasets/shivamb/netflix-shows

[4] Mayer-Schönberger, V. and Cukier, K., 2013. Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt.

[5] J. Geraghty, "How pop culture is influencing baby names," The Conversation, 2020. [Online]. Available: https://theconversation.com/how-pop-culture-is-influencing-baby-names-144562

[6] Australian Bureau of Statistics, "Births, Australia," 2023. [Online]. Available: https://www.abs.gov.au/statistics/people/population/births-australia