

Gender Classification in Modern Society

Gabrielle Beinars

DSC680-T301 Applied Data Science (2221-1)

<https://github.com/gbeinars>

Which Domain?

The domain from which the dataset was obtained is two population-based surveys which examines demographics and characteristics of adults that identify as nonbinary, including ages 18-60.

Majority of data science projects that currently exist with the purpose of classifying gender, only focus on binary targets of male and female. This is an incorrect generalization in today's society and culture, and reinforces the idea that there are only two categories, although this is no longer the case. Within data science, if something is not included in the project, it allows for the assumption that it does not exist, for example, with gender classification as male or female.³

1 Legal Recognition of Non-Binary Gender (background info),

https://en.wikipedia.org/wiki/Legal_recognition_of_non-binary_gender

2 News Discussion on Third Gender Option (background info),

<https://www.usatoday.com/story/news/2017/06/21/third-gender-option-non-binary/359260001/>

3 A Primer on Non-Binary Gender and Big Data (topics current stand in Data Science info),

<https://civic.mit.edu/index.html%3Fp=1165.html>

4 Data Sharing for Demographic Research, TransPop, US, 2016-2018, (download link)

<https://www.icpsr.umich.edu/web/DSDR/studies/37938/datadocumentation>

5 TransPop US Transgender Population Health Survey, Methodology and Technical Notes (study info)

<https://static1.squarespace.com/static/55958472e4b0af241ecac34f/t/5ef1331ecc152d3b60f068d6/1592865567882/TransPop+Survey+Methods+v18+FINAL+copy.pdf>

6 Male, Female, Nonbinary: The Ethical Responsibility of the IT Community (background info)

<https://er.educause.edu/articles/2019/3/male-female-nonbinary-the-ethical-responsibility-of-the-it-community>

7 Legal Recognition of Non-Binary Gender in US (background info),

https://en.wikipedia.org/wiki/Legal_recognition_of_non-binary_gender#United_States

8 1.2 million nonbinary people live in the US, a new study says (background info),

<https://www.cnn.com/2021/06/23/us/nonbinary-survey-study-number-trnd-wellness/index.html>

9 Nonbinary LGBTQ Adults in the US, Generations and TransPop Study Results (study info),

<https://williamsinstitute.law.ucla.edu/publications/nonbinary-lgbtq-adults-us/>

10 US Demographic Data (study info), [https://usafacts.org/state-of-the-](https://usafacts.org/state-of-the-union/population/?utm_source=bing&utm_medium=cpc&utm_campaign=ND-StatsData&msclkid=0c140a1892791bcaa87b6b2b0b6c94dc)

[union/population/?utm_source=bing&utm_medium=cpc&utm_campaign=ND-StatsData&msclkid=0c140a1892791bcaa87b6b2b0b6c94dc](https://usafacts.org/state-of-the-union/population/?utm_source=bing&utm_medium=cpc&utm_campaign=ND-StatsData&msclkid=0c140a1892791bcaa87b6b2b0b6c94dc)

11 A Groundbreaking Study Reveals There Are Over One Million Nonbinary Adults in the US (background info), <https://www.them.us/story/study-reveals-over-one-million-nonbinary-adults-united-states>

12 Transgender, Non-Binary, Intersex: Support & Activism Club (background info on terminology), <https://studentweb.bellevuecollege.edu/transactivism/non-binarygenderqueer/gender-terminology/>

Which Data?

The dataset that I'll use for this project is included above as citation 4 – The Williams Institute at the UCLA School of Law conducted a survey-based study to examine the demographic data to estimate the population of non-binary adults in the US. The data was collected from 2016 to 2018 and consists of various variables, including, age, race, census region, sexual identity, sex at birth, gender identity, household income, personal income, education, urbanicity, and more.

The data download link (citation 4) includes multiple files, cisgender codebook, transpop codebook, manifest, related literature, citations, and questionnaires used.

Research Questions? Benefits? Why analyze these data?

I plan to begin with basic exploratory data analysis approach by examining each applicable variable, cleaning the data (including any necessary imputation), and exploring relationships between variables. Various visualizations will be created to better understand and evaluate the variables included in the dataset. I plan to develop an understanding of the demographic parameters included in the dataset, and lastly, create a model to predict identity.

The following questions will be answered at the completion of this project:

- What portion of the population identifies as nonbinary?
- Are there any significant factors that better predict gender identity?
- How well does the classification model predict gender identity?

The results of this project will provide insight to modern-day demographics of gender classification.

What Method?

Using common data science EDA libraries, including pandas, numpy, and matplotlib, EDA will be conducted and any required data cleaning will be completed. Scikit-learn will be my library of choice for machine learning and creating a classification prediction model.

Potential Issues?

There are a couple potential challenges I foresee with this project. I have some concern for the lack of accurate data on those that classify as nonbinary. In the past, many may have shied away from honesty due to societal fears and judgement. While nonbinary people have always been a part of our population, it was not always recognized. Another obstacle I plan to discuss heavily in the results are ethical implications of applying a gender classification project. Gender is not as easily defined as it may appear, and I expect that the nonbinary category will be in the minority in any gender dataset. It is a complicated topic that can encourage many discussions.

Concluding Remarks

Within modern society, gender has been a widely discussed topic, with an additional category emerging, in addition to female and male, nonbinary. Multiple countries have become to legally recognize a third

gender classification, referred to as nonbinary, and defined as those born with sex characteristics that “do not fit the typical definitions for male or female bodies”.¹ On a federal level within the US, in February of 2020, a legislation was introduced that would add the third classification option to US passport applicants, and this is expected to be implemented by the end of 2021.

As this is a recent implementation for US history, I wanted to further explore gender classification, with the addition of a nonbinary category. Most gender classification projects focus on a binary response of male or female, but this is outdated in today’s time.

Oregon became the first state in the US to accept an option of nonbinary, X, on forms of state education. The non-profit organization that drove this change, Lambda Legal, stated “Presenting an identification document that does not accurately reflect one’s sex and is inconsistent with one’s gender identity can trigger invasions of privacy, prejudice, stigma, violence and discrimination and harassment in a wide variety of settings, including in employment, education, public accommodations, health care, housing and interactions with the government, including with law enforcement.”²