

# Gender and Higher Education: An Analysis of IPEDS Data

Cori N. Miller

Northwest Missouri State University, Maryville MO 64468, USA  
s545767@nwmissouri.edu

**Abstract. Keywords:** data analytics · higher education

## 1 Introduction

I would like to explore the Higher Education Domain. This is my current field, and there is a lot of data available to work with. I will likely use a combination of data sets from the National Center for Education Statistics, IPEDS, and institutional data from the college where I work. Specifically, I would like to look at equity gaps in higher education as they relate to gender and using gender to predict higher education outcomes. The steps to completing this project are defining the problem, collecting and preparing the data, building and evaluating a model, and deploying the model. To address important questions about gender equity in higher education, I will use high-quality data from IPEDS and NCES, appropriate data modeling techniques, and rigorous evaluation of model performance. Because the focus of this research has only one component, gender, it may be subject to contextual and data limitations.

### 1.1 Goals of this Research

Specify exactly your aims of this paper. Also, write a sentence how you will address for each individual goal.

## 2 Data and Methods

For this project, I am using a custom data set I created using the National Center for Education Statistics (NCES) data tools. This is a public tool that uses IPEDS surveys to create meaningful data about higher education. The Integrated Post-secondary Education Data System (IPEDS) is a set of surveys conducted annually by the U.S. Department of Education's NCES.

I created this data set using the Graduation Rates Survey. It includes a custom list of comparison institutions that I chose based on their similarity to my own institution, number of students by gender, and the graduation rate by gender. For this project, I am focusing on the impact of gender ratio on

higher education outcomes. I will be using the graduation rates by gender and population ratios as the major data attributes.

The data set has 298 records and 14 fields of structured data in CSV format. It is mostly numeric data, and since I was able to customize the list of institutions and variables, the data is very high quality. It does not have any missing fields or formatting issues. I used a data tool to make a custom file so I did not have to use any web scraping techniques.

### 3 Results

Presentation of the findings from the data modeling

### 4 Conclusion

Summary of the key findings

□

### References

1. <https://nces.ed.gov/ipeds/>
2. <https://nces.ed.gov/>
3. O'Connor, P.: Why is it so difficult to reduce gender inequality in male-dominated higher educational organizations? a feminist institutional perspective. *Interdisciplinary Science Reviews* (2020)