

Project Title: Gender Difference Indicators and Share of female STEM graduates

Team Names & Emails: Sara Clemente (saraclem@andrew.cmu.edu), Sharon John (sharonjo@andrew.cmu.edu), Gracie Siu (gsiu@andrew.cmu.edu)

Project Description:

1. Why is this topic relevant?

Gender disparities persist in STEM fields worldwide, with women being underrepresented in these areas compared to men. Addressing these disparities is crucial for achieving gender equality and ensuring that all individuals have equal opportunities to pursue careers in high-demand fields.

2. Who does this topic affect? Where does it happen? When did it happen?

This topic affects various stakeholders, including individuals, communities, industries, and economies. For individuals, it impacts career opportunities, earning potential, and professional fulfillment. Communities benefit from diverse perspectives and solutions that arise from gender-inclusive STEM environments. Industries benefit from tapping into a broader talent pool, fostering innovation, and addressing skill shortages. Economies benefit from increased productivity, competitiveness, and economic growth when all segments of the population are fully engaged in STEM fields.

The scope: Gender disparities happen in STEM worldwide, although the extent of these disparities may vary across regions and countries. While some countries have made significant strides in promoting gender diversity in STEM, others continue to face significant challenges. This issue is not limited to a specific geographic location but rather spans across continents and cultures.

Timeline:

Gender disparities in STEM have historical roots, stemming from systemic biases, social norms, and cultural expectations. While progress has been made in recent decades to address these disparities, significant gaps persist. Understanding the historical context of gender inequalities in STEM is essential for designing effective interventions and policies to promote diversity and inclusion.

What are your motivations for addressing this topic?

As three women in STEM we know the beauty but also challenges in this field. Addressing gender disparities in STEM is not only a matter of social justice but also an economic imperative. By ensuring equal access and opportunities for all individuals in STEM fields, societies can unlock the full potential of their human capital, drive innovation, and build a more equitable and prosperous future for everyone.

Datasets:

1. Gender data: <https://genderdata.worldbank.org/>
 - Format: CSV
 - Metadata:

The Gender Data Portal is more than just a database of sex-disaggregated data and gender statistics. Data sources include United Nations, UNESCAP, Eurostat, World Economic Forum, WHO, IMF, Institute for Democracy and Electoral Assistance, and more (World Bank).
 - Variables: Multiple indicators representing gender differences by Country, Year. Indicators span assets, legal rights, education rights and completion, employment, life expectancy, GDP.
2. Share of female STEM graduates:
<https://genderdata.worldbank.org/indicators/se-ter-grad-fe-zs/?fieldOfStudy=Science%2C%20Technology%2C%20Engineering%20and%20Mathematics%20%28STEM%29>
 - Format: CSV
 - Metadata:

For aggregate data, each economy is classified based on the classification of World Bank Group's fiscal year 2020 (July 1, 2019-June 30, 2020) (Worldbank)

Statistical concept and methodology: Percentage of female graduates by field of study in tertiary education is calculated by dividing the number of female graduates in a given field of education from tertiary education by the total number of graduates in the same field, and multiplying by 100.

Data on education are collected by the UNESCO Institute for Statistics from official responses to its annual education survey. All the data are mapped to the International Standard Classification of Education (ISCED)

to ensure the comparability of education programs at the international level. The current version was formally adopted by UNESCO Member States in 2011. The reference years reflect the school year for which the data are presented. In some countries the school year spans two calendar years (for example, from September 2010 to June 2011); in these cases the reference year refers to the year in which the school year ended (2011 in the example). Source: UNESCO Institute for Statistics (<http://uis.unesco.org/>). Data as of March 2020.

- Variables: Share of graduates by field, female (%). Broken down by Country, Year.

Project Questions:

1. In the coming years, are women in certain countries more likely to graduate with a STEM major than others?
 - Target variable(s) from data set: 'Share of STEM Graduates', 'Country Name'
 - Task type: prediction
2. What factor(s) influence women's access to STEM fields?
 - Target variable(s) from data set: Feature(s) in prediction model
 - Task type: prediction
3. Can we expect the proportion of women entering STEM fields to grow in the future? If yes, what factors will influence this growth?
 - Target variable(s) from data set: 'Share of STEM Graduates', 'Year', 'Country Name'
 - Task type: prediction

How to run the file:

All the information for our sprint is in one notebook. Reading the CSV, cleaning the datasets, merging and 3 visualizations

1. Correlation Matrix
2. Box plots
3. Line plot

References:

- <https://pandas.pydata.org/docs/reference/api/pandas.melt.html>

- Lab 2 from EDA
- https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.pivot_table.html
- <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.astype.html>
- <https://genderdata.worldbank.org/resources/?tab=gender-data-portals>

Submission

- Include your Jupyter Notebook/s in the repo
- Add your dataset to your .gitignore if you are working on it in your local repo, but make sure to provide an exact link for us to download your dataset/s
- Modify the second half of your README from Sprint 2 to include your project description, questions, etc. Here's a sample: [Sprint3_README_Sample-1.md](#)
- [Download Sprint3_README_Sample-1.md](#)
- **Push everything to the same repo you had from Sprint 2 before Thursday, February 29th at 11:59 PM**

NOTES:

Gender data:

<https://genderdata.worldbank.org/>

Share of female STEM graduates:

<https://genderdata.worldbank.org/indicators/se-ter-grad-fe-zs/?fieldOfStudy=Science%2C%20Technology%2C%20Engineering%20and%20Mathematics%20%28STEM%29>