BUDT704 Project Proposal:

EmpowHer: Analyzing Women’s Labor Market

Group 2 - The Pay-Checkers

Shamit Kotak, Andreas Mayr, Siddhant Virkar, Ishita Joshi, Asheer Mogal, Yue Zhao

# **Introduction:**

Globally, around 50% of all women are in paid employment, while the equivalent figure for men is 80%. When women work, they usually earn less, even in the societies where women already have high levels of employment and are perhaps more highly educated than men.

The goal of our project is to understand why levels of employment and earnings differ between women and men and why it is important for socio economic development of a region as well. The issue relates to the most efficient use of society’s resources. It is economically inefficient for jobs not to go to the most qualified person and, if pay differs for performing the same work, women may be disincentivised to work and have a career.

## **Questions of Interest**:

* What’s the relation between education level and employment rate gap?
* Are birth rates correlated to unemployment rate (by education level)?
* What are the differences between paternity and maternity leave in the United States?
* What is the correlation between GDP and women’s labor participation?
* How does women’s labor participation rate change over time?
* How much of the global labor rate gap is due to cultural norm differences?

*All of these questions can be analyzed by country or globally*

# Data Processing and Analysis

## **Dataset Description**

1. The main dataset we will be using comes from the **World Bank Gender Data Portal**. We found it online on the World Bank website, downloadable as a CSV. It has 305,545 rows and 57 columns. This is an aggregate dataset consisting of up to 1153 distinct quantitative and qualitative (boolean) variables for 265 countries and/or regions.
2. The **Parental Leave Policies dataset** contains a csv table with 1,601 records, one for each company. Each record contains the company's name & industry, as well as crowdsourced information on the paid & unpaid weeks off they offer as part of both their maternity & paternity leave policies (when available).
3. The (World Bank) **Employment and Time** **dataset** examines gender differences in labor market involvement, demonstrating that women hold many vulnerable, low-paying professions and are less likely than males to participate in formal employment. It focuses on **important indicators** such as **legal rights, workplace safety, rates of child labor, sexual harassment laws, and employment distribution across industries**.

## **Data Processing Tasks**

* Scrape data from web pages:
  + We will download the data from The World Bank Gender Data Portal and kaggle
* Indexing, selection, and filtering
  + We will be filtering the data to only include data from 1993 and forward, giving us data for the last 30 years.
  + We will select different groups of data concerning different questions to do the relevant analysis.
* Managing missing data
  + Managing missing data involves two primary strategies: indexing missing values for transparency and future reference, or deleting them for simplicity and streamlined analysis. The choice between methods depends on the dataset's nature and the impact of missing values on the analysis, ensuring accurate and reliable results.
* Data transformation
  + We may transform the qualitative data to quantitative data when it’s needed.
  + We may change the data type to make calculations easier when it’s needed.
* Merging and reshaping data
  + We will check on the data within each group and pick reasonable variables to compile them.

## **Data Analysis**

We will generate an array of line charts, bar graphs, scatter plots, and various other visual representations. Most of these visualizations will be static, but we can also include interactive elements for country-based filtering. Our dataset will also enable the creation of a global heatmap representing different metrics.

* We'll employ both a scatter plot, complete with a trendline, and a line graph featuring global averages to illustrate GDP and labor rates.
* To analyze correlations, we'll utilize both correlation matrices for a clear representation of correlations and standard plots to help us visualize the data, especially if it doesn't exhibit the expected correlations.
* In the context of maternity and paternity leave, we'll use a pie chart where the entire pie represents a full year. This will make it simple to observe the disparities in allowed paid and unpaid time between men and women.
* Additionally, we'll employ a Time Series Plot to visualize data points at consecutive time intervals.

## **Expected Findings**

1. What's the connection between educational attainment and the employment rate?

* The findings might reveal that, in recent years, women's educational levels have been steadily increasing, but disparities in employment and earnings persist, both at lower and higher education levels. The data may indicate that the extent of these disparities hasn't changed as much as anticipated.

2. Are birth rates correlated to unemployment rate (by education level)?

* Higher education levels (such as bachelor's degrees or advanced degrees) can result in a negative association between birth rates and unemployment rates. This may indicate that people with higher levels of education tend to put off starting a family when the economy is uncertain and may give priority to their careers or their financial security.

3. How do paternity and maternity leave policies in the United States differ?

* Discrepancies in perceptions and practices of maternity and paternity leave can worsen the gender pay gap. Advocating for equal paternity leave can challenge traditional norms, encourage shared parental responsibilities, and contribute to narrowing this gap.

4. How does women’s labor participation rate change over time?

Expected results about changes in women's labor force participation rates over time can take many different forms and differ depending on particular geographic areas, economic circumstances, and social factors. Possible discoveries and their implications are listed below:

* Over the past few decades, women's labor force participation rates have likely increased generally in many industrialized nations. This shows that more women are joining the profession due to a variety of reasons, including shifting social standards, easier access to education, and a desire for financial security.
* Results may indicate that women who have completed more education are more likely to work. This illustrates the growing significance of education in gaining access to improved employment prospects.

5. What is the correlation between GDP and women’s labor participation?

* We might see a U-shaped association. Due to their need for additional income, women may have higher labor force participation rates in less developed economies. Women's involvement rates may decline as economies develop because of increased access to education and other opportunities. Women's labor force involvement may, however, increase once more as economies achieve higher levels of development, frequently in occupations requiring greater ability and professionalism.

6. What portion of the global gender employment gap can be attributed to differences in cultural norms?

* Cultural norms influence societal expectations, individual choices, and employer practices, collectively shaping the global gender employment gap. Addressing these cultural barriers is vital to achieving gender equality in labor markets worldwide.

# **Project Timeline**

| **Task** | **Task Lead** | **Due Date** |
| --- | --- | --- |
| Data Collection [Web Scraping] | Shaunak | 19th Oct |
| Data Filtering | Shaunak | 24th Oct |
| Data Selection | Shamit | 25th Oct |
| Handling missing data | Ishita | 28th Oct |
| Data Transformation and Merging | Siddhant | 31st Oct |
| Data Cleaning | Mia | 6th Nov |
| Data Processing | Asheer | 9th Nov |
| Data Analysis | Siddhant | 15th Nov |
| Data Analysis - Educational Attainment vs. Employment Rate | Mia | 18th Nov |
| Data Analysis - Birth Rates vs. Unemployment Rates | Andreas | 18th Nov |
| Data Analysis - Comparison of Paternity and Maternity Leave Policies | Asheer | 18th Nov |
| Data Analysis - Global Gender Employment Gap and Cultural Norms | Shamit | 18th Nov |
| Data Analysis - GDP vs. Women's Labor Participation | Shaunak | 20th Nov |
| Data Analysis - Trends in Women's Labor Participation | Mia | 20th Nov |
| Data Analysis - Correlation Analysis - Constructing correlation matrices to identify relationships between various socio-economic factors and gender employment disparities. | Andreas | 20th Nov |
| Comparison of Employment Disparities at Different Education Levels | Asheer | 22th Nov |
| Predicting Future Trends | Siddhant | 22th Nov |
| Inference from all Processing and Analysis | Ishita | 25th Nov |
| Completion of final analysis and observations | Andreas | 25th Nov |
| Document Verification and Submission of Assignment | Shamit | 1st Dec |