

An Analysis of the Global Gender Inequality in the Workforce

Shreya Pillamari, Harshini Dinesh, Samantha Pierre, Diya Kumar, Shivani Rajan

Northeastern University, Boston, MA, USA

Abstract

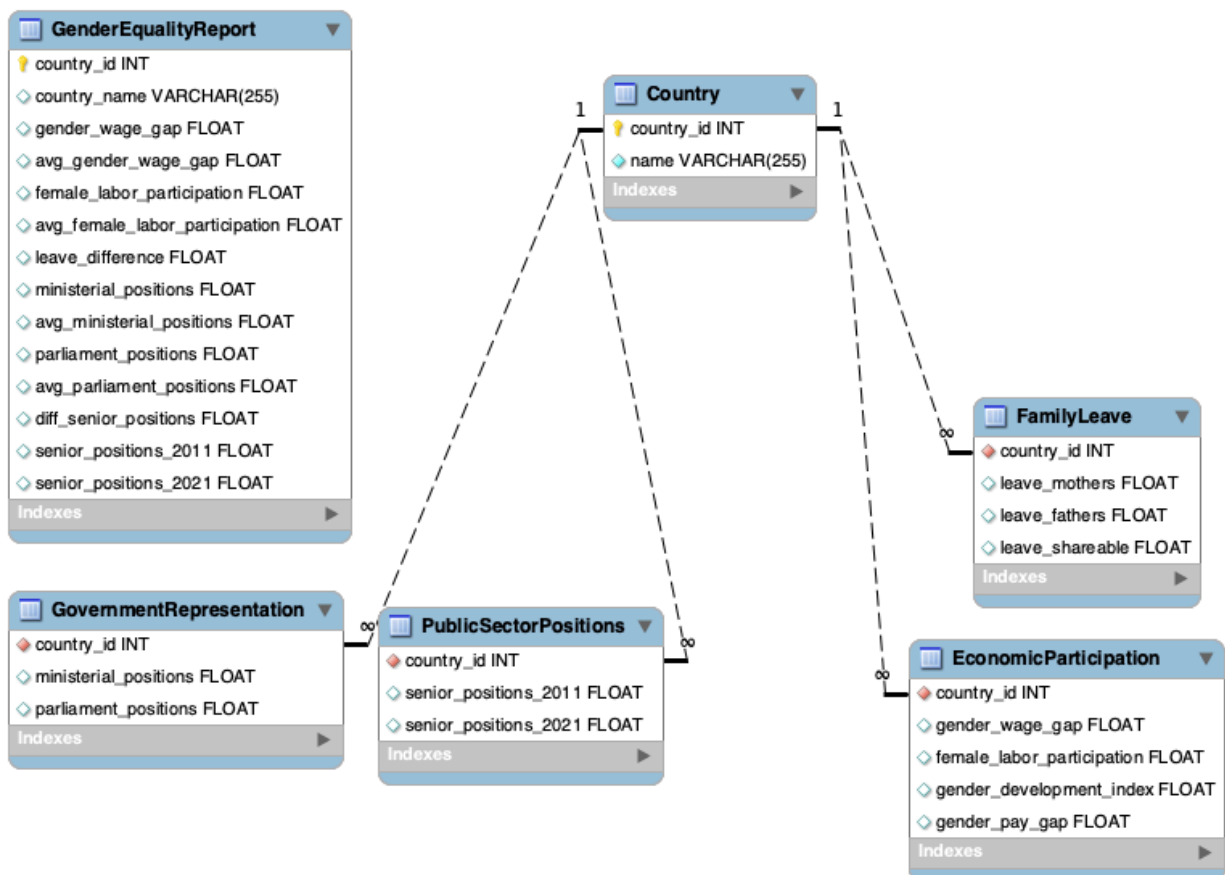
This project investigates gender inequality in the workplace through an SQL database application, and it analyzes gender development metrics across various countries. It is meant to answer questions about representation, economic participation, and family leave disparities. It uses interconnected tables and looks at gender wage gaps, labor force participation, leadership representation, and policy inequities. You can query specific metrics, gender equality reports, and compare nations too. There are database procedures, test cases, and triggers to help compare the different tables.

Introduction

Gender inequality is a huge issue globally, and it influences economic, political, and social areas. This project explores the factors that contribute to these problems, such as workforce representation, economic participation, and family leave policies. By creating the `gender_development_database`, we have a tool that helps look at trends and insights that will help inform the general public.

This database takes data from many different types of gender inequality- public sector leadership, parliamentary roles, gender pay gaps, and labor force participation rates. It has queries to provide comparisons between men and women, and tools to create country specific reports. The database also uses automated updates with triggers to keep data accurate and consistent. The database shows us the connection between policy, economics, and representation for gender inequality, helping us better understand and address the issue.

Database Design



Our database, `gender_development_database`, is used to analyze data related to gender development metrics across different countries. The `country` table is the central entity in this database- each country is uniquely identified by `country_id` and `name`. This table is the reference point for all the other tables, creating a one-to-many relationship. The table `FamilyLeave` has information about family leave policies, with `leave_mothers`, `leave_fathers`, and `leave_shareable`. The table `PublicSectorPositions` focuses on gender representation in public sector roles- and uses two timeframes (2011 and 2021) to provide analysis overtime of public sector roles. The table `GovernmentRepresentation` tracks women's participation in government positions. The table for `EconomicParticipation` provides information on a few economic standards: gender wage gap, female labor force participation, gender development index, and gender pay gap. Every single one of these tables has a relationship with the `country` table, and each table contains a foreign key, `country_id`, which references the primary key in the `country` table. This ensures that all data in the related tables are connected to specific countries. The `gender_equality_report` table generates reports on gender equality for specific countries, including metrics like `gender_wage_gap`, `female_labor` participation, and calculated values like `leave_difference`. It references the `Country` table through `country_id` and provides insights for comparing companies. Essentially, this table helps analyze gender development metrics.

Data Sources and Methods

To comprehensively analyze global gender inequality in the workforce, we gathered and processed data from multiple reputable sources, which made sure we had a strong base for our research. Our primary dataset included information on variables such as women's representation

in public sector leadership, government roles, economic participation, and family leave policies across various countries. This data was sourced from international databases, such as World Bank, UN Women, and OECD. To maintain the reliability of our data, we prioritized that it be from within the last five years. Before breaking down our data in MySQL, we gathered as much qualitative insights into gender development indices, pay gaps, labor force participation rates, and family leave durations. To ensure accuracy we cross referenced the data against official publications and reports. Data discrepancies or missing entries were also noted for potential imputation or exclusion. This data was broken down into a few subcategories (country information, public sector representation, government representation, economic indicators, and family leave entitlements) that became cleaned and structured data which was exported as a CSV file, and served as the input for the backend database. The data generated to this process of manual web scraping and inserting into a was compiled and then placed into a SQL backend file to generate the database with the data loaded. The backend of our project was primarily built using MySQL. We broke down the data into easily approachable variables such as Country, which stores the country ID and name of the country, PublicSectorPositions, which tracks the percentage of women in senior public sector positions, GovernmentRepresentation, the variable that records the percentage of women in ministerial and parliamentary positions, EconomicParticipation, which follows the values of paid family leave for mothers, fathers, and shareable leave. To populate the database, we used structured INSERT statements based on the preprocessed data. With values that we couldn't generate, we initialized them with NULL because we still needed those spots to be filled in some way. This way of creating the backend of the project ensures that the database is scalable and able to handle queries and analysis.

User Cases (Application Prototype) or Analysis (Data Science)

We created our SQL database with the ability for users to ask the following questions regarding our gender inequality data:

1. Which countries had lower than 30% of women in senior positions in the public sector in both 2011 and 2021?
2. Which countries had a lower percentage of women in parliament than the percentage of women in ministerial positions?
3. What were the top 5 countries with the largest gender pay gap?
4. Which countries had a female labor force participation below 50%?
5. In which countries did the share of women in senior positions in the public sector, increase by at least 10 percentage points between 2011 and 2021?
6. Which countries had their share of women in senior positions in the public sector in 2021 exceed 40%, while the gender pay gap was greater than 15%?
7. In which countries did the duration of paid leave earmarked for mothers be at least five times the duration of paid leave earmarked for fathers?
8. In which countries did women hold less than 20% of ministerial positions while the gender wage gap exceeded 10%?
9. In which countries did the share of women in senior positions (2021) exceed 50% while the Gender Development Index was below 1?
10. Can you group countries by GEO Region, calculate the average share of women in senior positions in the public sector for 2011 and 2021, and identify regions with the largest and smallest improvements over time?

11. Which countries have women making up at least 50% of senior positions in 2021 but their labor force participation is less than 60%?
12. Can you rank countries by the combined average of the share of women in senior positions (2021), women in ministerial positions, and women in parliament?
13. In which countries do the share of women in senior positions (2021) exceed 40%, and the shareable paid family leave weeks most often taken by mothers exceeds 80%.?
14. What is the inequity score for each country and can you rank them to show the top 5 countries with the greatest inequities?

To answer these questions, we used a series of join, group by, ordering, and aggregation queries. Users can also use our database to ask other questions as well. We also created two procedures so that users can retrieve the information they need from our database. Our first procedure called 'gender_equality_report' was created so that when a user inputs a specific country name, they can get a report on gender equality in that specific country in relation to the database's average data. For example, if a user inputted 'Korea' as a country that they wanted a gender equality report on, the procedure would run and generate a report with its country_id, country_name, gender_wage_gap, avg_gender_wage_gap, female_labor_participation, avg_female_labor_participation, and more.

To ensure that this procedure runs smoothly we also created two triggers that recalculates average values (i.e. average gender wage gap, average women in ministerial positions) when a country updates its specific value (i.e. Korea's gender wage gap, Korea's number of women in ministerial positions). Finally, we created another procedure called 'countriesfamilyleaveforboth' which identifies whether a user-inputted country gives equal or unequal days in family leave for

both mothers and fathers. This helps users get data and compare countries with large versus small disparities or gender inequalities. We also created multiple test cases to see if both procedures work accurately and efficiently.

	country_id	country_name	gender_wage_gap	avg_gender_wage_gap	female_labor_participation	avg_female_labor_participation	leave_difference	ministerial_positions	avg_ministerial_positions
▶	2	Austria	22.2	11.1273	46.71	54.7632	11.67	26	34.5036
	21	Korea	17.3	11.1273	54.25	54.7632	10.8571	21	34.5036
	30	Portugal	7.7	11.1273	51.33	54.7632	1	44.1	34.5036
	33	Spain	5.2	11.1273	57.46	54.7632	0	61.4	34.5036

parliament_positions	avg_parliament_positions	diff_senior_positions	senior_positions_2011	senior_positions_2021
41	33.575	9.3	31.2	40.5
27.2	33.575	NULL	NULL	NULL
35.7	33.575	11.8	39	50.8
45.6	33.575	3.2	39.9	43.1

Conclusions

This project was primarily fueled by our cumulative interest in gender differences in the workforce and we wanted to come to an understanding of what data existed out there to help us solve questions related to gender wage differences. Through this project, we developed a SQL database that analyzed gender inequality variables in various countries. We further highlighted patterns and trends through the use of backend SQL, queries, and ER diagrams. Overall, this project is a step toward addressing gender inequality by raising awareness of disparities.

Author Contributions

Diya Kumar: Created the ER Diagram, helped find data sources, wrote the Introduction, Abstract, and Database Design part of the project report and the slideshow.

Shreya Pillamari: Created the backend script creating the categories and filling the insert statements, helped find data sources, wrote half the data sources and methods and the conclusion

Shivani Rajan: Created the ER Diagram, helped find data sources, wrote the Introduction, Abstract, and Database Design section of the project report, and contributed to the slideshow.

Harshini Dinesh: Created the two procedures, two triggers, and multiple test cases in SQL and edited the queries to not have any errors. Wrote the ‘User Cases’ part of the project report and made slide 5. Found the OECD website where we got our initial data from, and added data to the spreadsheet.

Samantha Pierre: Created the SQL Questions that we used to construct our queries, created queries 1-14, hosted multiple zoom meetings for project check-ins, added comments to the SQL, helped in making the presentation (slide 6) and helped in looking for data sources.

References:

1. “Gender Equality and Work.” *OECD*, www.oecd.org/en/topics/gender-equality-and-work.html. Accessed 25 Nov. 2024.