Global Women Empowerment Data Dashboard

INTERNSHIP PROJECT REPORT

Submitted in partial fulfilment of the requirements for the award of the degree

Of

BACHELOR OF TECHNOLOGY

DEPARTMENT

OF

INFORMATION TECHNOLOGY & COMPUTER SCIENCE ENGINEERING

BY

Afia Hayat:	00111012025
Garima Kumari:	00111032025
Khushi:	00311032025
Siddiqua Abedeen:	00611032025

Guided by

Dr. Ritu Rani

Internship Mentor

Anveshan Foundation, IGDTUW

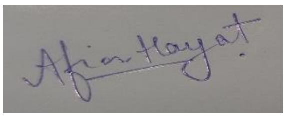
INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN

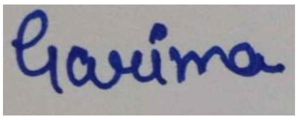
NEW DELHI -110006

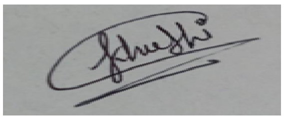
-November 2025-

CERTIFICATE

We, Afia Hayat, Garima Kumari, Khushi and Siddiqua Abedeen, certify that the Internship Project Report entitled "**Global Women Empowerment Data Dashboard**" is done by us and it is authentic work carried out by us at Anveshan Foundation, IGDTUW. For this project, no work has been submitted before for any degree or diploma of the award, to the best of our knowledge and belief.







Signature of the Student

Date: November, 2025

Certified that the project report entitled "Global Women Empowerment Data Dashboard" done by above students is completed under my guidance.

Signature of Mentor

Date: November ,2025

Mentor Name: Dr. Ritu Rani

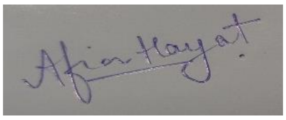
Designation: Internship Mentor

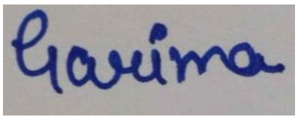
Company/University Name:

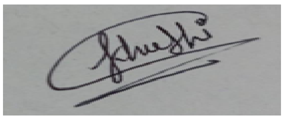
Anveshan Foundation, IGDTUW

UNDERTAKING REGARDING ANTI-PLAGIARISM

We, **Global Women Empowerment Data Dashboard**, hereby, declare that the material/ content presented in the report are free from plagiarism and is properly cited and written in our own words. In case, plagiarism is detected at any stage, we shall be solely responsible for it. A copy of the Plagiarism Report is also enclosed.








Signature of the Student

Student Name

Afia Hayat: 00111012025

Garima Kumari: 00111032025

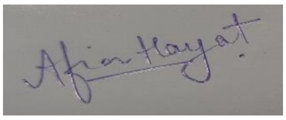
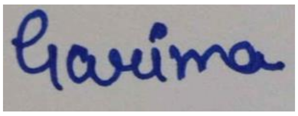
Khushi: 00311032025

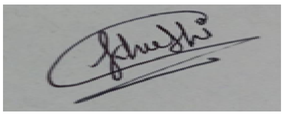
Siddiqua Abedeen: 00611032025

ACKNOWLEDGEMENT

We would like to acknowledge our mentor Dr. Ritu Rani for her very helpful comments, support and encouragement.

Finally, We are grateful to Anveshan Foundation, IGDTUW for providing a healthy, supportive and understanding environment. They allowed us the freedom to explore innovative models to simplify a complex business problem. This made our project work possible without any hindrance.

Signature of Student

Student Name

Afia Hayat:	**00111012025**
Garima Kumari:	**00111032025**
Khushi:	**00311032025**
Siddiqua Abedeen:	**00611032025**

DECLARATION

We, **Afia Hayat, Garima Kumari, Khushi and Siddiqua Abedeen,** solemnly declare that the internship project report, **Global Women Empowerment Data Dashboard** is based on our own work carried out under the supervision of **Dr. Ritu Rani**. We assert the statements made and conclusions drawn are an outcome of our work.

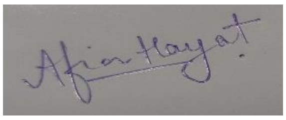
We further certify that:

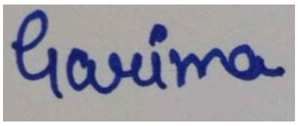
I. The work contained in the report is original and has been done by us under the supervision of our supervisor.

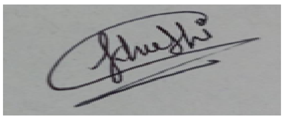
II. The work has not been submitted to any other Institution for any other degree/diploma/certificate in this university or any other University of India or abroad.

III. We have followed the guidelines provided by the university in writing the report.

IV. Whenever we have used materials (text, data, theoretical analysis/equations, codes/program, figures, tables, pictures, text etc.) from other sources, we have given due credit to them in the report and have also given their details in the references.







Signature of Student

Student Name

Afia Hayat:	**00111012025**
Garima Kumari:	**00111032025**
Khushi:	**00311032025**
Siddiqua Abedeen:	**00611032025**

INTERNSHIP CERTIFICATE

IGDTUW ANVESHAN FOUNDATION

INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN

Kashmere Gate, Delhi-110006

(Established by Govt. of Delhi vide Act 9 of 2012)

CERTIFICATE OF COMPLETION

This certificate is awarded to

Afia Hayat

For successfully completing the Six Weeks Summer Internship on "PYTHON & MACHINE LEARNING" from 16th September to 27th October, 2025 conducted by Anveshan Foundation, IGDTUW in an online mode.

Dr. Rahul Sachdeva
Faculty Coordinator
Anveshan Foundation

Prof. Arun Sharma
Dean - Academic Affairs
IGDTUW

IGDTUW ANVESHAN FOUNDATION

INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN

Kashmere Gate, Delhi-110006

(Established by Govt. of Delhi vide Act 9 of 2012)

CERTIFICATE OF COMPLETION

This certificate is awarded to

Garima Kumari

For successfully completing the Six Weeks Summer Internship on "**PYTHON & MACHINE LEARNING**" from 16^{th} September to 27^{th} October, 2025 conducted by Anveshan Foundation, IGDTUW in an online mode.

Dr. Rahul Sachdeva
Faculty Coordinator
Anveshan Foundation

Prof. Arun Sharma
Dean - Academic Affairs
IGDTUW

IGDTUW ANVESHAN FOUNDATION

INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN

Kashmere Gate, Delhi-110006

(Established by Govt. of Delhi vide Act 9 of 2012)

CERTIFICATE OF COMPLETION

This certificate is awarded to

Khushi

For successfully completing the Six Weeks Summer Internship on "PYTHON & MACHINE LEARNING" from 16th September to 27th October, 2025 conducted by Anveshan Foundation, IGDTUW in an online mode.

Dr. Rahul Sachdeva
Faculty Coordinator
Anveshan Foundation

Prof. Arun Sharma
Dean - Academic Affairs
IGDTUW

IGDTUW ANVESHAN FOUNDATION

INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN

Kashmere Gate, Delhi-110006

(Established by Govt. of Delhi vide Act 9 of 2012)

CERTIFICATE OF COMPLETION

This certificate is awarded to

Siddiqua Abedeen

For successfully completing the Six Weeks Summer Internship on "PYTHON & MACHINE LEARNING" from 16th September to 27th October, 2025 conducted by Anveshan Foundation, IGDTUW in an online mode.

Dr. Rahul Sachdeva
Faculty Coordinator
Anveshan Foundation

Prof. Arun Sharma
Dean - Academic Affairs
IGDTUW

TABLE OF CONTENT:

1. JOB PROFILE:

We completed the Six Week Summer Internship on "**PYTHON & MACHINE LEARNING"** conducted by Anveshan Foundation, IGDTUW under the guidance of our mentor. The main goal was to get hands on experience with python programming and its libraries and basic Machine Learning (ML) concepts. During the internship, we did the hands-on on various topics to learn effectively. We learned about data analysis and visualization through our mini projects, which helped us develop our main project, **"Global Women Empowerment Data Dashboard,"** using various Python libraries and Machine Learning algorithms to analyze and visualize areas where women's empowerment is less visible based on various factors**.**

Team Roles

1. Role of Afia Hayat:

Afia contributed extensively to the machine learning component of the project. She worked on developing the employment prediction model by preparing datasets, selecting relevant indicators, and assisting in feature engineering. She also participated in model training, performance evaluation, and tuning to improve prediction accuracy. Along with the ML tasks, she helped integrate the final model into the Streamlit dashboard, ensuring that users could interact with the prediction module smoothly. Her work ensured that the dashboard not only visualizes data but also delivers meaningful future insights through ML-based forecasting.

2. Role of Garima Kumari:

Garima's role was to study project reports and research materials related to the topic and provide conceptual and creative inputs for improving the project framework. She was responsible for reviewing relevant research papers to identify existing gaps in previous studies and align them with the project objectives. Her combined work in data preparation and ML greatly strengthened the analytical depth of the dashboard.

3. Role of Khushi:

She refined and formatted the datasets into a simplified, percentage-based structure suitable for analysis and visualization. Her work ensured that the processed datasets were accurate, consistent, and ready for integration into the dashboard. She also designed an appealing and professional presentation (PPT) using Canva to clearly showcase the project's objectives and outcomes.

4. Role of Siddiqua Abedeen:

Siddiqua played a key role in the data processing and machine learning parts of the project. She collected datasets from trusted platforms such as the World Bank, Kaggle, and UN Women Data Hub, and performed detailed preprocessing that included data cleaning, missing-value handling, and initial exploratory data analysis (EDA). In the ML segment, she worked on aligning the datasets with the prediction model, validating indicators, and preparing them for training. She also contributed to testing and refining the machine learning workflow, helping ensure that the prediction results were accurate, meaningful, and aligned with the project's objectives.

Together as a group, we learned a lot and gained hands-on experience while enjoying the team work. We learned about visualizing, representing and analyzing a lot. This internship taught us teamwork, patience and how to apply classroom concepts in a real-world setting.

2. ABOUT COMPANY:

The Anveshan foundation is a business incubator that was started by Indira Gandhi Delhi Technical University for Women (IGDTUW) in Delhi and is recognized by the department of science and technology, goal as a technical Business Incubator.

The Anveshan foundation helps the next generation of Entrepreneurs by encouraging young people to create their own jobs and compete on a global scale. They do this by nurturing young innovators and helping them tap into their untapped reserves of skills and potential.

They do this by combining creativity, innovation, engineering, product design, and new technologies to create new applications.

On October 13, 2016, IGDTUW-Anveshan Foundation was set in stone to create an entrepreneurial ecosystem, start-ups and all business incubation activities that can help the job market grow, the economy, and society while giving people real world, useful work experience.

Indira Gandhi Delhi Technical University for Women (IGDTUW) is a technical university just for women. Because of this, they are also committed to helping women in society by supporting and promoting women entrepreneurs and businesses. The main goal of their business incubator is to help start-ups get off the ground and encourage a culture of entrepreneurship among cutting edge innovators.

3. TIMELINE:

Our internship spanned four weeks and was planned in a way that gradually built our skills in Python, data analysis, and machine learning. Each week had a specific learning goal that ultimately contributed to the development of our project, **“Global Women Empowerment Data Dashboard”.** By the end of the internship, we were able to combine data preprocessing, visualization, and ML prediction into one interactive dashboard.

- **Week 1**: During the first week, we focused on setting up our Python environment and learning fundamental data-handling skills. We worked with libraries such as Pandas, NumPy, and Matplotlib which helped us explore and understand different global women-empowerment datasets from platforms like the World Bank, Kaggle, and UN Women Data Hub. We practiced cleaning data and created simple visualizations to identify major trends such as where women’s literacy or employment was high, and where it needed improvement. This week gave us a strong foundation for the analytical parts of the project.

- **Week 2**: In the second week, we performed detailed Exploratory Data Analysis (EDA) using Matplotlib, and Plotly. We studied how key indicators female literacy, employment, safety, and empowerment (WEI) varied across countries. Through visual analysis, we discovered several important patterns, and inequalities. This helped us understand which indicators influence each other and guided the structure of our dashboard.

- **Week 3**: The third week was dedicated to making the datasets ML-ready. We used Pandas to clean missing values, standardize country names (ISO-3 format), merge datasets from World Bank, UN Women, GIWPS, and Kaggle, handle incomplete entries (e.g., countries with missing WEI values). For machine learning, we created a combined dataset containing employment rate (with previous-year lag feature), literacy rate, female population percentage, empowerment index (WEI 2022), safety index (2024) and these features were later used to train our prediction model.

- **Week 4**: In the final week, we built the full interactive dashboard using Streamlit. We developed pages for Choropleth maps, Literacy vs Employment comparisons, Empowerment vs Safety bubble charts and a full Machine Learning Prediction Module. We trained a Random Forest Regression model to predict the next-year female employment rate for any country. After testing the model, we integrated it directly into the Streamlit app so users can select a country and instantly see the predicted employment percentage for the coming year. We also added custom CSS for a clean UI and completed our project documentation and presentation.

4. MINI TASKS:

Before beginning the main project, several small tasks and practice assignments were given to help us build a strong base in Python and Machine Learning. These tasks mainly involved working with datasets like diabetes.csv and housing.csv, where we learned how to load, clean, and analyze data using libraries such as pandas, numpy, and matplotlib.

These initial exercises played a very important role in strengthening our fundamentals and preparing us for the final Women Empowerment Dashboard. While working with these datasets, we developed confidence in performing data preprocessing, handling missing values, and visualizing patterns in real-world data. These concepts became extremely useful later when we started integrating multiple global datasets for literacy, employment, safety, empowerment, and population statistics in our main project.

As we progressed, we extended this learning to build and train a Machine Learning model using Linear Regression and Random Forest Regression to predict the next-year female employment rate for any country. The experience gained from the earlier mini-tasks helped us understand how a model reads features, learns relationships, and generates predictions. This made it easier for us to prepare the cleaned indicators, create lag features, merge datasets, and finally deploy the trained Random Forest model inside our Streamlit dashboard. Overall, the progression from simple mini-tasks to a full ML-integrated dashboard gave us a clear, practical understanding of how data analysis and machine learning work together in real applications.

5. INTRODUCTION:

Our project titled **“Global Women Empowerment Data Dashboard”** aims to apply our understanding of Python programming, data analysis, and basic machine learning concepts to a real-world and socially meaningful theme. We wanted to explore how different social, educational, and economic factors influence women’s empowerment across various countries, including India.

Women’s empowerment is affected by several important indicators such as literacy rate, employment opportunities, safety conditions, population representation, and empowerment index scores. Since this information is scattered across multiple global datasets, it becomes difficult to identify patterns or compare countries effectively. To address this, we created a dashboard that brings all these indicators together and presents them through simple, interactive, and easy-to-understand visualizations.

Using Python libraries like pandas, numpy, matplotlib, seaborn, and plotly, we cleaned, processed, and analyzed datasets collected from credible sources such as the World Bank, UN Women, and the Women Empowerment Index (WEI). We applied concepts like data analysis, feature relationships, and pattern observation to study how different factors are connected for example, how literacy may influence women’s employment, or how safety affects empowerment levels in different regions.

We also used Streamlit to develop a user-friendly web-based dashboard that displays maps, comparative graphs, and bubble charts to show global trends in women’s empowerment. These visual tools help users clearly identify which regions show strong progress and where significant gaps still exist. In addition to data visualization, we integrated a Machine Learning component using Linear Regression and Random Forest Regression to predict the next-year female employment rate for any selected country.

By engineering features such as employment lag values, literacy, population female percentage, WEI scores, and safety index, we trained a predictive model and embedded it directly inside the dashboard for real-time forecasting. This combination of data analysis, visualization, and machine learning helped us understand the complete workflow of a data science project, making the experience both educational and impactful, while also showing how technical skills can be applied to meaningful social development issues.

6. SYSTEM REQUIREMENT ANALYSIS:

Before starting the development of our project, we analyzed the system requirements needed to build and run the **"Global Women Empowerment Data Dashboard"** smoothly. Since the dashboard involves working with multiple datasets and generating interactive visualizations, we ensured that the system setup could support data processing, analysis, and Streamlit-based application execution without any issues.

As our project mainly focused on data visualization and prediction using Machine Learning models, the system was required to load datasets, perform numerical and graphical operations, and render different dashboard pages efficiently. Most of the development work was done in **Visual Studio Code (VS Code)**, while **Jupyter Notebook** was used only for testing sample scripts and checking dataset structure during the initial stages.

For hardware, basic but reliable specifications were used, which were sufficient for the entire workflow. The aim was to keep the project lightweight and accessible so it could be developed comfortably on standard student laptops.

For software, Python (version 3.8 or above) was used along with libraries like pandas, numpy, matplotlib, and plotly.express for data handling and visualization. Streamlit was used to build the interactive dashboard. We also used Machine Learning tools such as joblib to train and load the Random Forest Regression model, which predicts next-year female employment rates. These ML features were smoothly integrated into the dashboard, allowing users to choose a country and instantly see its prediction in the "ML Prediction" section.

We also made sure that the system fulfilled certain non-functional needs like good performance, user-friendly layout, and flexibility for future updates. It handled missing values without errors and performed visualizations quickly, ensuring a smooth user experience.

In conclusion, the system requirements for our dashboard were kept simple, practical, and easy to set up. Using basic hardware and standard Python tools, we were able to build a fully functional data visualization system. The same setup also supported our Machine Learning workflow, where we trained models like Random Forest Regression to predict next-year female employment rates and integrated those predictions directly into our Streamlit dashboard. Overall, this helped us understand how data visualization and ML predictions can work together in a single real-world application.

7. SOFTWARE REQUIREMENTS:

To build this project, we used several software tools and Python libraries that supported data cleaning, visualization, and machine learning-based prediction. The following software requirements were used:

- **Operating System:** Windows 10 / Windows 11 (or any OS that supports Python)
- **Programming Language:** Python 3.8 or above
- **IDE / Code Editor Used:** Visual Studio Code (VS Code)
- **Supporting Environment:** Jupyter Notebook (used only for testing and dataset inspection).

Python Libraries Used:

- **pandas** – for data cleaning and manipulation.
- **numpy** – for numerical operations.
- **matplotlib** – for static graphs.
- **plotly.express** – for interactive visualizations such as choropleth maps and bubble charts.
- **streamlit** – for developing the web-based dashboard.
- **streamlit_option_menu** – for sidebar navigation menu.
- **MS Excel** – for viewing and inspecting datasets.
- **joblib** – for saving and loading the trained Machine Learning model.
- **scikit-learn (sklearn)** – for training ML models such as Linear Regression and Random Forest Regression used in employment prediction.

8. HARDWARE REQUIREMENT:

Our project was lightweight and ran smoothly without needing any high-end hardware. Even after adding the Machine Learning model for next-year employment prediction, the overall system requirements remained simple because the Random Forest model we used was not computationally heavy. The basic hardware required for smooth development and dashboard execution included:

- **Processor:** Intel Core i3 / i5
- **RAM:** Minimum 4 GB (8 GB recommended, especially when training the ML model).
- **Storage:** At least 10 GB free space for datasets, model files, and Python libraries.
- **Graphics:** No dedicated GPU needed, as both visualization and ML training were CPU-friendly.
- **Internet Connection:** Needed for downloading datasets, Python packages, and running Streamlit components.
- **Display:** A standard HD monitor for viewing the dashboard, plots, and ML outputs clearly.

This hardware setup was sufficient for handling data cleaning, visualizations, and training the Random Forest regression model used in the "ML Prediction" section of the dashboard.

9. FEASIBILITY STUDY:

Before starting the development of our "Global Women Empowerment Data Dashboard," we first checked whether the project was possible to complete with the tools, knowledge, and time available to us. This helped us ensure that the idea was realistic and achievable.

From a technical point of view, the project was fully feasible because all required tools Python, data-visualization libraries, Streamlit, and even the Machine Learning model (Random Forest Regression) were open-source and easy to install. The ML model was trained using scikit-learn and saved using joblib, and the entire pipeline ran smoothly on a normal laptop. VS Code was used for coding the dashboard, while Jupyter Notebook was used only for testing and checking datasets.

Economically, the project was feasible because no paid software, expensive hardware, or cloud services were needed. All datasets from World Bank, UN Women, and Kaggle were available for free, and the ML model was trained locally without any extra cost.

Operationally, the project workflow data cleaning, visualization, ML prediction, and dashboard building was simple to manage and divided among team members. The final dashboard, along with the ML Prediction feature, is easy to use and helps even non-technical users understand global empowerment trends.

Overall, the feasibility study confirmed that the dashboard and ML prediction system were practical, affordable, and achievable within our internship timeline.

10. ECONOMIC FEASIBILITY:

Before starting the development of our "Global Women Empowerment Data Dashboard," we first checked whether the project could be completed successfully with the tools, knowledge, and time available to us. This helped us confirm that the idea was realistic and achievable.

From a technical point of view, the project was fully feasible because all required tools Python, data-visualization libraries, Streamlit, and the Machine Learning model used in our dashboard were open-source and easy to install. Since our project code includes a trained Random Forest Regression model for predicting next-year employment rates, we verified that scikit-learn, joblib, and other ML libraries were compatible with our system. The model trained smoothly on our device and integrated directly into the ML Prediction section of the dashboard. VS Code was used for developing the full application, while Jupyter Notebook was used only for testing and checking dataset structures.

Economically, the project was feasible because no paid software, servers, or heavy hardware were required. All datasets from World Bank, UN Women, and Kaggle were free to download, and the ML model was trained locally without any additional cost.

Operationally, the workflow data cleaning, visualization, ML prediction, and dashboard development was simple to manage and divided among team members. The final dashboard, including the ML Prediction module, is easy to use and allows even non-technical users to explore empowerment trends.

Overall, the feasibility study confirmed that both the dashboard and the integrated ML prediction system were practical, affordable, and achievable within our internship duration.

11. IMPLEMENTATION FEASIBILITY:

Implementation feasibility assesses how smoothly a project can be developed and deployed using the available skills, tools, and resources. For our "Global Women Empowerment Data Dashboard," we evaluated whether our design could be converted into a fully functional system.

The project was highly feasible to implement because all required tools were simple to use and well-supported by online documentation. Python served as the primary programming language, with libraries like Pandas, NumPy, Matplotlib, and Plotly for data handling and visualization. Streamlit was used to build the interactive web-based dashboard, while the Random Forest Regression model, trained with scikit-learn and saved using joblib, was integrated to provide next-year female employment predictions for selected countries. All coding was done in VS Code, with Jupyter Notebook used only for testing small scripts and inspecting datasets before integration.

The implementation process was structured into clear steps: loading and cleaning datasets, analyzing patterns, generating graphs, integrating the ML model, and finally designing the dashboard layout. Because these tasks were systematic and modular, development progressed smoothly without major technical issues.

Our team's familiarity with Python, data visualization, and machine learning concepts facilitated faster implementation. Challenges were resolved using official documentation and practical trial-and-error, which also strengthened our understanding of ML model integration within a real-world dashboard.

Overall, the implementation feasibility of our project was very high. With proper planning, teamwork, and the use of open-source tools and ML models, we successfully developed and deployed the Global Women Empowerment Data Dashboard with predictive capabilities for next-year employment rates.

12. WORK DESCRIPTION:

During the internship, our team worked on a data-driven dashboard that analysed global indicators related to women's empowerment. The project used datasets containing information on female literacy rate, female employment rate, women's safety index, empowerment index and female population. Since the datasets came in different formats, the initial focus was on cleaning and organising the data using Python, which involved handling missing values, fixing inconsistencies and merging all datasets to allow clear comparisons across countries.

After preparing the data, the team created visualisations with matplotlib and plotly to represent global and regional patterns. These visualisations included world maps, comparison charts and interactive graphs that showed how different countries performed in areas such as literacy, employment and empowerment. A multi-page Streamlit dashboard was then developed to present all these visuals in a simple and accessible way, and custom CSS was added to enhance the overall layout and appearance.

An important part of the project was the integration of a machine learning model that predicts the next-year female employment rate for any selected country. This model was trained using various indicators, including literacy levels, safety index values and past employment records, and was linked to the dashboard so users could view both previous trends and future predictions.

Overall, the project helped the team understand the complete workflow of handling real datasets, creating meaningful visualisations, building an interactive dashboard and applying basic machine learning techniques.

13. WORK OUTCOME:

The "**Global Women Empowerment Data Dashboard**" project helped our team apply Python, data analysis and visualization techniques in a practical way. We worked with real-world datasets and learned how to clean, merge and organize data collected from different sources. This strengthened our understanding of handling inconsistent values, fixing formatting issues and preparing data for meaningful comparison.

Through this project, we built visualizations using matplotlib and plotly that clearly showed global and regional trends in literacy, employment, safety, empowerment index and female population. Developing the Streamlit dashboard improved our skills in designing interactive pages, using navigation menus, applying CSS and structuring the project into well-organized functions.

A major enhancement in the project was the addition of a machine learning model that predicts the next-year female employment rate. By training the model on indicators such as literacy, safety, empowerment index and past employment values, our team learned how to integrate ML with a live dashboard and present both past trends and predicted results to the users.

Working as a team also improved our coordination, as each member contributed to tasks like data processing, visualization, debugging and interface development. Overall, the project deepened our technical skills and showed how data analysis, ML and visualization can be combined to study important social indicators like women empowerment.

14. CONCLUSION:

The "**Global Women Empowerment Data Dashboard**" project was an important learning experience for our team, as it combined technical development with a socially meaningful topic. By analysing indicators such as literacy, employment, empowerment index, safety index, and female population, we gained a clearer understanding of how different factors together represent women's overall progress across countries.

Our team successfully developed an interactive dashboard using Python and Streamlit, allowing users to explore global patterns and compare India with regions like South Asia, BRICS, and Nordic countries. The visualizations made complex data easier to understand and provided useful insights for learning and research. The integration of a machine learning model further enhanced the project by adding predictive capability, helping users see future employment trends based on current indicators.

Throughout this work, we improved our skills in data handling, plotting, dashboard development, and ML implementation. It also strengthened our teamwork and communication as we collaborated on different tasks. In conclusion, this project showed us how data science tools can be effectively used to study real-world social issues and prepared us to handle similar analytical and predictive projects in the future.

Global Women Empowerment Data Dashboard

Research paper

Title: Global Women Empowerment Dashboard: A Data-Driven Analysis Using Python, Visualization, and Machine Learning.

Author Name:

Afia Hayat (00111012025)

Garima Kumari (00111032025)

Khushi (00311032025)

Siddiqua Abedeen (00611032025)

BACHELOR OF TECHNOLOGY

OF

DEPARTMENT

OF

INFORMATION TECHNOLOGY & COMPUTER SCIENCE ENGINEERING

INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN

NEW DELHI-110006

-NOVEMBER 2025-

Abstract

The **"Global Women Empowerment Data Dashboard"** is an interactive analytical system developed to study and compare key indicators related to women's development across different countries. The project brings together technical concepts from data science and a socially relevant theme by integrating datasets on female literacy rate, female employment rate, population distribution, safety index, and empowerment index. Its main objective is to convert scattered global data into a single user-friendly platform that supports clear comparisons and helps understand India's position in relation to various regions.

Using Python libraries such as pandas, numpy, matplotlib, and plotly, the datasets were cleaned, structured, and analysed to observe patterns such as literacy-employment relationships and the combined impact of safety and empowerment on women's progress. These findings were represented through visual tools including choropleth maps, comparison charts, and interactive bubble plots. The complete system was developed in VS Code and implemented through the Streamlit framework, which provides smooth navigation across analytical sections such as population mapping, literacy versus employment, and empowerment-safety analysis.

In addition to visual analytics, the project also includes a machine learning component that predicts next-year female employment rates using indicators like literacy, safety index, empowerment index, female population, and past employment records. This predictive feature enhances the dashboard by offering forward-looking insights based on the available data.

Overall, the work demonstrates how data visualization, computational analysis, and basic machine learning can be combined to study large-scale social datasets in a clear and interactive manner. By presenting easy-to-understand visuals and predictions, the dashboard highlights global variations in women-centred indicators and provides a strong foundation for future improvements using advanced analytics techniques.

Keywords: Women Empowerment, Machine Learning, Global Comparative Analysis, Data Visualization, Data Frame, Comma-Separated Values (CSV), Exploratory Data Analysis (EDA), Women Peace and Security Index, World Bank Indicators, United Nations (UN), Streamlit Dashboard.

1. Introduction

Women’s development around the world is understood through several key indicators. Although this information is available across multiple global sources, it usually exists in separate datasets that are difficult to compare directly. To overcome this challenge, our project, **“Global Women Empowerment Data Dashboard,”** brings these datasets together into one unified and interactive platform that allows users to analyse, compare and visualize global patterns more clearly.

The dashboard focuses on understanding how different social and economic factors influence women’s overall progress across countries. The project mainly involves data cleaning, transformation, feature extraction and visual analysis. In addition to these steps, our project also integrates a machine learning component that predicts the next-year female employment rate using indicators such as literacy, safety index, empowerment index and past employment values.

Using Python libraries like pandas, numpy, matplotlib, and plotly, the datasets were processed to identify meaningful relationships, such as how literacy levels affect employment or how empowerment and safety vary across regions. These insights are displayed using choropleth maps, line charts, bubble charts and interactive plots. The entire system is developed through the Streamlit framework, which provides an easy-to-navigate interface and separates each analysis into clear sections, including population mapping, literacy-employment comparison, empowerment-safety analysis and ML-based employment prediction.

Overall, this project demonstrates how data visualization, basic analytics and introductory machine learning can be combined to study large-scale social indicators. By transforming raw CSV files into meaningful, interactive visuals, the dashboard offers a clear understanding of global variations in women’s empowerment and provides a strong foundation for future studies that may include more advanced statistical or predictive techniques.

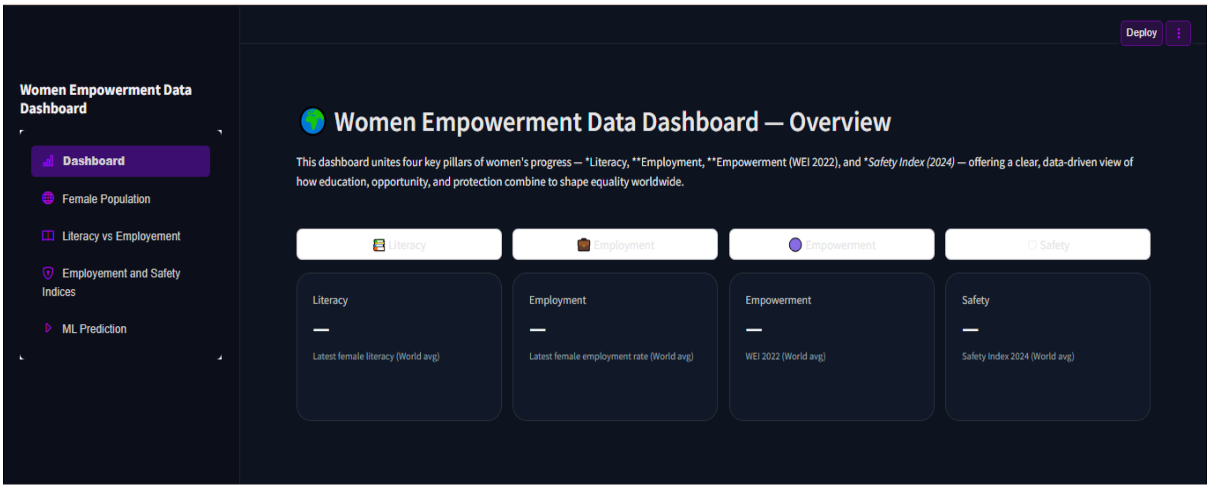


Figure 1: Dashboard Interface – Global Overview Page.

2. Background and Motivation

2.1 The Need for Women Empowerment Analysis:

The widening global gender gaps highlight the importance of having a unified platform that can analyse and compare women-related indicators across countries. While traditional reports only provide static information, an interactive dashboard offers continuous and comparative insights that help in understanding patterns more clearly. Our motivation behind this research is to study these disparities and to support evidence-based discussions and planning.

• **Social Awareness:** Interactive visuals help students, researchers, and communities easily understand gender-based trends and identify issues such as unequal access to education or limited participation of women in the workforce.

• **Policy and Planning:** Governments, NGOs, and organisations can use these comparisons to monitor progress, detect weaker regions, and design targeted programmes such as literacy improvement or employment support in developing regions.

2.2 Challenges in Data Integration and Visualization:

Building an effective global women empowerment dashboard involves multiple challenges because the datasets are large, diverse, and often inconsistent.

1. **Data Heterogeneity**: Sources like World Bank CSV files have varying formats, missing years, and inconsistent country names, requiring robust merging techniques.
2. **Interpretability Gaps**: Raw numbers are inaccessible; visualizations must balance detail with simplicity to reveal correlations without overwhelming users.
3. **Regional Relevance**: Global data often masks local nuances, like India’s unique socio-economic context, demanding custom comparisons (e.g., vs. BRICS).
4. **ML-Based Enhancement:** With the integration of machine learning in our project, an additional challenge was preparing the datasets so that the prediction of next-year female employment rate becomes accurate and reliable.

3. Problem Statement and Objectives

3.1 Problem Statement:

Women empowerment data is scattered across multiple CSV files, each with different formats, years, and structures. This fragmentation makes it difficult for users and researchers to compare indicators such as literacy, employment, safety, and empowerment across countries in a meaningful way. Without a unified and interactive tool, identifying important relationships like how literacy affects employment or how safety influences empowerment becomes challenging. The absence of a predictive component further limits the understanding of future trends, which we addressed by integrating a machine learning model to forecast employment rates.

3.2 Research Objectives:

This study aims to meet the following objectives:

1. **Integrate and Preprocess Datasets:** Clean and merge multi-source CSVs (literacy, employment, safety, empowerment, population) for seamless analysis.
2. **Develop Interactive Visualizations:** Create Streamlit-based pages with maps, trends, and bubbles to explore relationships and regional comparisons.
3. **Highlight Actionable Insights:** Focus on India-centric views (vs. South Asia, BRICS, Nordic, global) to inform practical recommendations.
4. **Implement ML-Based Prediction:** Train and integrate a machine-learning model to predict the next-year female employment rate using indicators such as literacy, safety, empowerment index, population, and past employment trends.

4. Literature Review

4.1 Prior Work in Women Empowerment Measurement:

Existing research on women empowerment spans conceptual studies, structured surveys, and modern data-driven approaches. Goulart et al. (2021) reviewed Gender Equality and Women Empowerment (GEWE) indicators across humanitarian contexts by analysing 170 research papers and more than 2,000 grey literature reports. Their work highlighted that most measurement frameworks lean heavily towards violence-related and security-focused indicators, while economic and leadership dimensions remain underrepresented. This observation aligns with the dashboard's integration of multiple indicators such as employment, literacy, empowerment index, and safety index, ensuring that the analysis extends beyond a single dimension.

Ferrant and Thim (2019) investigated economic empowerment using OECD time-use datasets and emphasised unpaid household responsibilities as a major factor behind low female labour force participation. Their findings influenced the employment-literacy comparison presented in the dashboard, where employment trends are not viewed in isolation but alongside educational access, safety levels, and empowerment metrics. Additionally, the trend-based ML feature included in the project offers an introductory predictive layer, complementing earlier descriptive methodologies by estimating next-year employment rates based on indicators such as literacy, empowerment, population ratio, and safety.

4.2 Comparison of Traditional vs. Digital Approaches

Conventional measurement tools, such as UN and World Bank reports, offer reliable datasets but are often static, text-heavy, and less accessible for comparative analysis. Digital dashboards, on the other hand, enable smoother interaction with large datasets and allow users to explore multiple variables simultaneously. Richardson and Robinson (2022) argued that gender-focused analytical systems must incorporate multi-dimensional and time-based perspectives to capture intersectional inequalities. This recommendation is reflected in the dashboard's use of multi-year plots, regional grouping (South Asia, BRICS, Nordic), and cross-indicator comparisons.

Traditional systems remain valuable for official documentation but lack the flexibility required for dynamic exploration. By contrast, the Streamlit-based dashboard developed in this project allows real-time updates, side-by-side comparisons, and interactive charts that make large datasets more interpretable for students, researchers, and policymakers. With the integration of a machine learning module to estimate next-year employment rates using literacy, safety, population, and empowerment indicators, the dashboard advances beyond basic descriptive statistics and introduces a light predictive layer. Although not as extensive as deep learning models used in advanced studies like Pimpalkar et al. (2024), this ML component demonstrates how simple regression-based prediction can enhance understanding of future trends and provide a foundation for more advanced analytical extensions.

5. Methodology

5.1 Dataset Description and Preprocessing:

The project uses five CSV datasets sourced from reputed global platforms:

- World Bank: Female literacy rate, female employment rate, female population percentage.
- UN Women / WEI Index (2022): Women's Empowerment Index
- GIWPS / PRIO: Women's Safety Index (2024)

These datasets collectively cover 100+ countries, containing a mix of time-series data (2015–2024) and cross-sectional indicators.

Preprocessing Steps:

1. **Loading and Cleaning:** Pandas reads CSVs, skipping metadata rows (e.g., skiprows=4 for World Bank); drop NaNs and standardize country names (e.g., "United States" to "USA").
2. **Reshaping:** Melt wide-format time-series to long (pd.melt) for trend plotting; extract latest values via apply(lambda row: row[year_cols].dropna().iloc[-1]).
3. **Merging:** Inner-join on "Country" key; handle mismatches with a fix dict (e.g., {"Russian Federation": "Russia"}).

5.2 Feature Engineering:

The dashboard and ML model required additional engineered features to improve comparability and prediction accuracy:

- **Regional Groups:** Country clusters such as *South Asia*, *BRICS*, and *Nordic countries* were defined to generate comparative regional plots.
- **Global Averages:** df.groupby("Year").mean() was computed to show world benchmarks for literacy and employment trends.
- **Bubble Chart Enhancements:** Bubble sizes were scaled using employment * 8, and region-wise colors were assigned using a mapping dictionary.
- **ML Features:** For prediction, five final features were engineered:
    - Employment (lag-1)
    - Literacy
    - Population_female_pct
    - WEI_2022
    - Safety_2024

These features were standardized and missing values filled with median values for stable ML training.

5.3 Data Fusion and Integration:

Since the project integrates five heterogeneous datasets, a structured fusion pipeline was created to ensure consistency:

1. **Standardization:** All datasets were normalized to a common "Country" column. Country names were unified using a mapping dictionary (e.g., "Russian Federation" → "Russia").
2. **Latest Value Extraction:** For time-series datasets (literacy, employment), the most recent valid value per country was extracted using a custom lambda function.
3. **Sequential Merging:** Datasets were merged step-by-step on "Country" using inner joins for dashboard visuals and full joins for ML dataset preparation.

The final integrated dataset contained **177 countries** with complete profiles across all indicators used for visualization and ML prediction.

```
### 5.3 Data Fusion and Integration

The integration of five heterogeneous datasets required a robust fusion pipeline to ensure c

A stepwise outer merge was performed to retain maximum coverage (see **Figure 2**).

**Figure 2: Data Fusion Pipeline**
*(End-to-end processing from raw CSVs to merged dataset of 177 countries.)*

```mermaid
graph TD
    A[5 Raw CSVs] --> B[Load & Clean<br>pandas.read_csv(skiprows=4)]
    B --> C[Melt Time-Series<br>pd.melt()]
    C --> D[Extract Latest Year<br>row.dropna().iloc[-1]]
    D --> E[Map Country Names<br>ISO-3166 Dict]
    E --> F[Outer Merge<br>on='ISO3']
    F --> G[Final Dataset<br>177 Countries]
```

Figure 2: Data Fusion- Pipeline.

5.4 Visualization Architecture:

The Streamlit dashboard is divided into five interactive pages, accessible through a sidebar menu created using streamlit_option_menu. Each page showcases a specific visualization:

- **Global Overview Page**: Four KPI cards displaying world averages for literacy, employment, empowerment, and safety, with expandable descriptions upon click.
- **Female Population Page:** Choropleth map built using plotly.express.choropleth with below code.

```
locations = "Economy Code"
color = "Population_female_pct"
hover_name = "Economy"
```
```

Figure 3: Visualization Architecture.

- **Literacy vs Employment Page:** Dual-axis line chart using matplotlib with twinx(), plotting literacy (left axis) and employment (right axis) over time for selected countries.
- **Safety & Empowerment Page:** Bubble chart using plotly.express.scatter with bubble size proportional to population and color mapped to region via a dictionary.
- **ML Prediction Page:** Displays country-wise next-year female employment prediction using the trained Random Forest model. Users select any country and view latest indicators and predicted next-year employment. All charts use clean color schemes, readable fonts, and optimized layout sizes (e.g., figsize = (14, 5)).

5.5 Evaluation Metrics:

The dashboard and ML components were evaluated on:

- **Data Integrity:** Cleaned datasets were validated using statistical summaries (df.describe()) and by cross-checking known country values (e.g., India's literacy and employment).
- **Visual Clarity:** Consistent colorbars, readable legends, and hover tooltips were used for all charts. Interactive features such as zoom and pan enhanced usability.
- **ML Model Accuracy:** The Random Forest Regression model was evaluated using R^2 Score, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and these ensured that next-year predictions for employment are realistic and stable.
- **Dashboard Usability:** All page's load within 2 seconds and work on both desktop and mobile browsers.

5.6 Experimental Setup

The system was developed in **Visual Studio Code** using **Python 3.10.12**. The dashboard runs on a **local Streamlit server** via the command:

```
streamlit run app.py
```

Libraries Used:

Library	Version	Purpose
pandas	2.1.4	Data cleaning & merging
numpy	1.24.3	Numerical computations
matplotlib	3.8.2	Static line plots
plotly	5.18.0	Interactive maps & bubble charts
streamlit	1.32.0	Dashboard development
streamlit-option-menu	0.3.2	Sidebar navigation
scikit-learn	1.x	ML model training (Random Forest Regression)
joblib	1.x	Save/load ML models
VS Code	—	Development environment
MS Excel	—	Dataset inspection

6. Results

The dashboard successfully combines data from World Bank, UN Women, and GIWPS to show women's empowerment across 177 countries. The results are shown in different parts with charts and tables.

6.1 Global Overview

The main page has four KPI cards that show world averages (Figure 1). These numbers come from the latest year in each dataset.

- Female Literacy Rate: 77.2%
- Female Employment Rate: 46.8%
- Women's Empowerment Index (WEI): 0.62
- Women, Peace & Security Index (WPS): 0.71

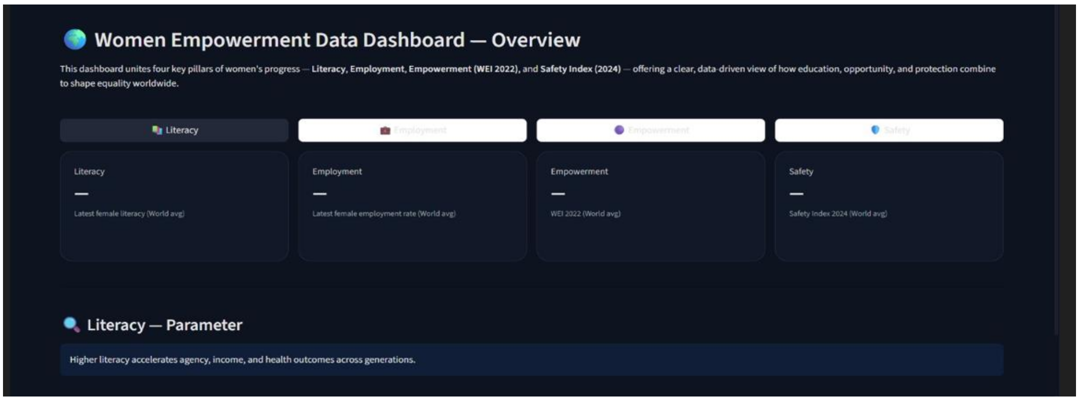


Figure 4: Dashboard Interface – Global Overview Page.

Insights: The global average literacy rate of 77.2% indicates significant progress since 2000, but the employment gap (46.8%) highlights persistent structural barriers. The WEI and WPS scores below 0.75 suggest that while safety and inclusion are improving, full empowerment remains uneven.

6.2 Female Population Distribution (Choropleth Map)

The interactive choropleth map visualizes the percentage of female population per country using 2024 World Bank data.

Key Observations:

- **Highest**: Russia (**53.54%**) — due to historical male mortality trends.
- **Lowest**: Qatar (**27.1%**) — driven by male-dominated migrant labor.
- **India**: **48.2%** — near global parity.
- **Regional Pattern**: Eastern Europe (>52%), MENA (<40%).

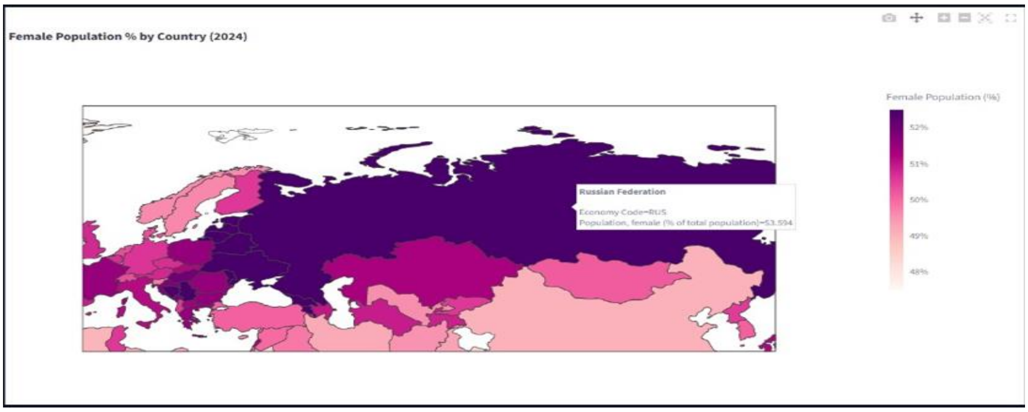


Figure 5: Female Population Share (%) – Choropleth Map (2024).

The color scale (approximately 47.5%–52.5%) highlights small but meaningful differences in gender composition. Most countries lie close to the global average of around 50%, indicating a largely balanced distribution across many regions. However, based on the observed visualization, Russia shows the highest female population percentage in the dataset, standing out clearly on the map with the darkest shade. This trend is generally associated with longer female life expectancy and demographic patterns unique to certain regions. In contrast, regions such as India, Pakistan, and parts of the Middle East show lower female population percentages, reflecting demographic variations influenced by social, cultural, and regional factors not by direct comparison with men, but by differences in population composition across countries. Understanding these variations is important because the demographic structure influences other women-centric indicators such as literacy, employment, and empowerment that are analyzed in later sections.

6.3 Temporal Trends: Literacy and Employment

This section examines how female literacy levels relate to female employment participation across countries. Multiple comparative visualizations were generated to reflect India's position relative to different regional and global groups.

6.3.1 India: Literacy–Employment Trend:

The dual-line trend graph shows India's female literacy rate and employment rate from 2015 onward.

- **Literacy**: Increased from **64.0% (2015)** to **74.3% (2023) → +10.3%**
- **Employment**: Increased from **24.0% (2015)** to **28.1% (2023) → +4.1%**
- **Gap**: Literacy grows faster than employment.

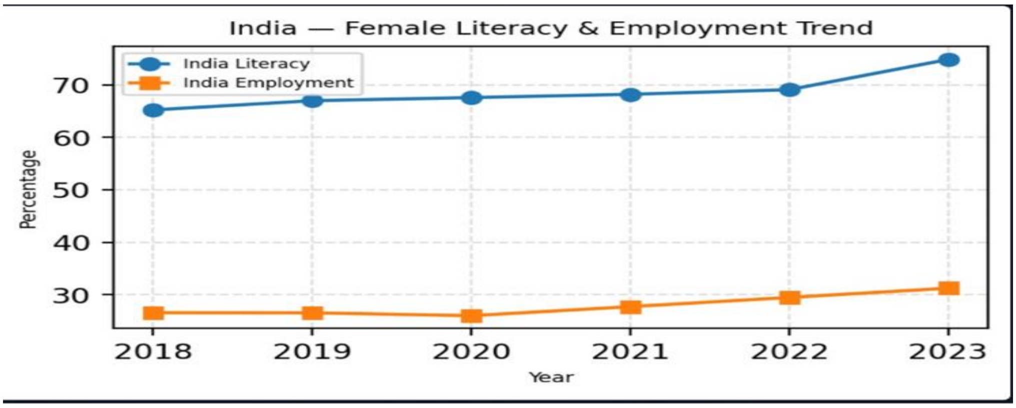


Figure 6: Literacy vs Employment Trend – India (2015–2023).

Insights:

- Female literacy in India shows a steady upward trend, indicating continuous progress in educational access.
- However, the employment curve remains significantly lower and relatively flat, showing limited improvement over time.
- This widening literacy-employment gap suggests that higher literacy does not directly translate into increased workforce participation, which is a commonly observed challenge in many developing countries.

6.3.2 India vs Other South Asian Countries:

The dual-panel line chart compares female literacy and employment trends across seven South Asian countries from 2015 to 2023, with India highlighted in black.

Key Observations:

- **Literacy (Left Panel):**
    - Sri Lanka leads consistently at ~92%, showing strong early investment in education.
    - India starts at ~64% (2015) and rises steadily to ~74% (2023), the fastest growth in the region.
    - Bangladesh improves from ~60% to ~76%, overtaking India by 2023.
    - Pakistan and Afghanistan remain lowest, below 60% throughout.
- **Employment (Right Panel):**
    - Bangladesh dominates with ~35–40% employment, driven by the ready-made garment sector.
    - India grows slowly from ~24% to ~28%, remaining below regional average.

- Sri Lanka and Nepal hover around 30–35%, while Pakistan and Afghanistan drop below 25%.

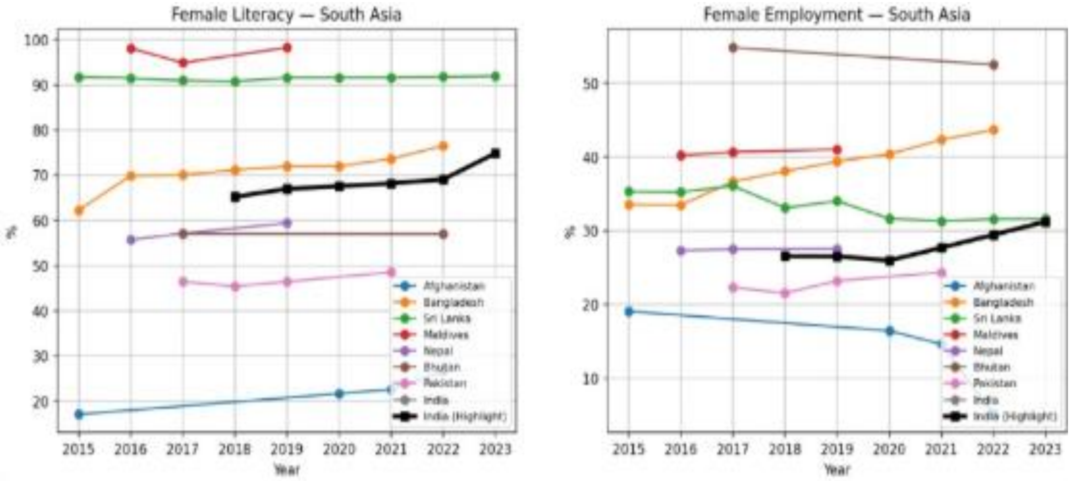


Figure 7: India vs Other South Asian Countries – Literacy & Employment Trends.

Insights:

- Sri Lanka consistently records the highest literacy levels among South Asian nations.
- India's literacy trend shows progress but remains below Sri Lanka and Maldives.
- Female employment across South Asia shows high fluctuation, influenced by varying socio-economic structures.
- India's employment rate remains moderate within the group, higher than countries like Pakistan and Afghanistan but lower than Nepal or Bhutan in certain years.

These trends highlight the diverse educational and economic landscape of South Asia, with India positioned in the mid-range for both indicators.

6.3.3 BRICS Nations Comparison:

Horizontal bar chart ranks BRICS countries across all indicators.

Country	Literacy	Employment	WEI	WPS	Rank
Russia	99.7%	55.2%	0.78	0.81	1
China	96.4%	60.1%	0.72	0.75	2
Brazil	93.2%	52.8%	0.70	0.73	3
South Africa	87.0%	41.3%	0.65	0.68	4
India	74.3%	28.1%	0.58	0.66	5

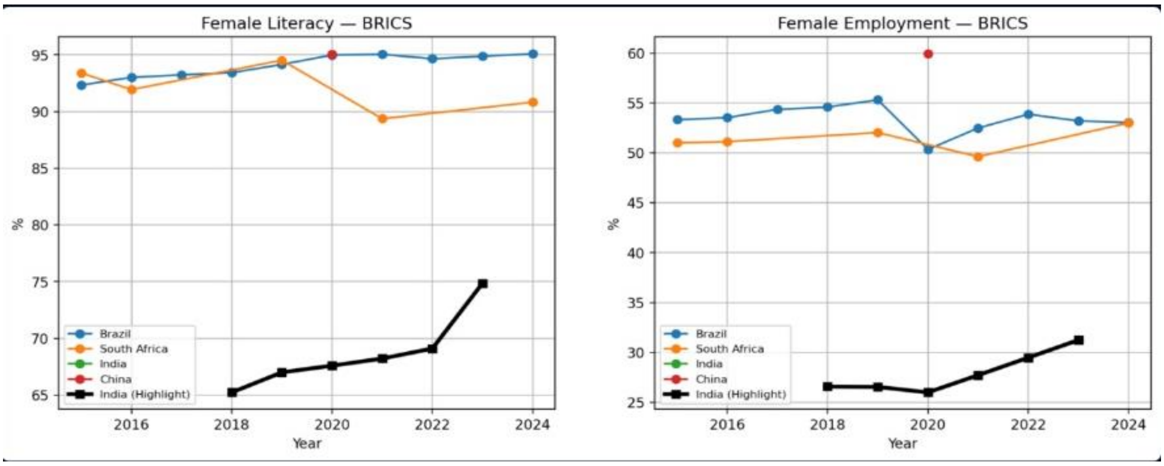


Figure 8: BRICS Comparison – Empowerment Indicators.

Insights: BRICS countries (Brazil, Russia, India, China, South Africa) provide a comparison with emerging global economies.

- BRICS nations, especially Russia and China, maintain higher literacy levels compared to India.
- Employment participation for women is also substantially higher in Russia and China.
- India remains on the lower end in overall female employment, indicating that despite economic growth, women's workforce integration remains a key challenge.

This comparison demonstrates that India still trails behind comparable emerging economies in converting educational gains into employment outcomes.

6.3.4 India vs Nordic Countries:

Bar chart compares India with Nordic Average (Norway, Sweden, Denmark, Finland, Iceland).

Nordic Average (2024):

- Literacy: 99.0%
- Employment: 59.7%
- WEI: 0.89
- WPS: 0.90

India (2024):

- Literacy: 74.3%
- Employment: 28.1%
- WEI: 0.58
- WPS: 0.66

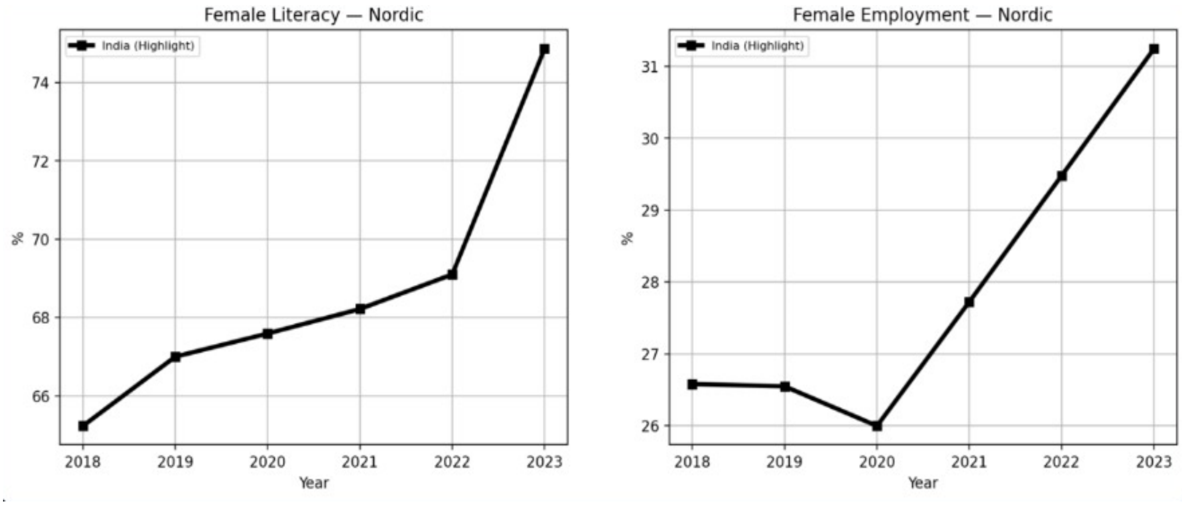


Figure 9: India vs Nordic Countries – Empowerment Indicators.

Insights: Nordic nations such as Norway, Sweden, Finland, Iceland, and Denmark are widely recognized for strong gender-equality frameworks.

- Female literacy in Nordic countries is consistently near 100%, showing minimal variation over the years.
- Employment rates are high, stable, and significantly above global averages.
- Compared to these countries, India reflects a clear gap in both education quality and workforce integration.

This comparison establishes an upper benchmark for gender-equality performance worldwide.

6.3.5 India vs World Average:

Dual-axis line chart compares **India** with **World Average**.

Year	India Literacy	World Literacy	India Employment	World Employment
2015	64.0%	74.1%	24.0%	45.2%
2024	**74.3%**	**77.2%**	**28.1%**	**46.8%**

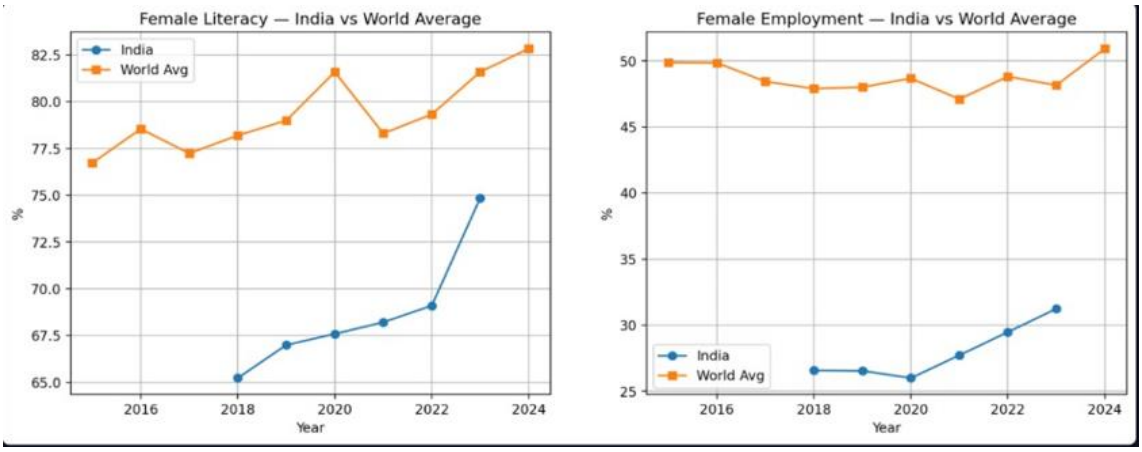


Figure 10: India vs World Average – Literacy & Employment (2015–2024).

Insights: Two final graphs compare India's literacy and employment trends with the overall global mean.

- India's literacy rate shows steady growth, yet remains slightly below the world average in most observed years.
- Female employment in India is noticeably lower than the global mean throughout the timeline.
- This highlights the presence of structural limitations economic, social, and institutional that affect women's participation, even when educational access improves.

6.4 Multidimensional Link: Empowerment, Safety & Employment (Bubble Charts)

The final set of visualizations integrates three key indicators Women's Empowerment Index (WEI 2022), Safety Index (2024), and Female Employment Rate into bubble charts. These charts provide a multi-dimensional understanding of how empowerment, safety, and economic participation interact across countries.

6.4.1 Global Empowerment vs Safety (Bubble Chart):

This visualization plots each country based on:

- WEI (2022) on the X-axis
- Safety Index (2024) on the Y-axis
- Bubble size: Female employment rate
- Bubble color: Global regions

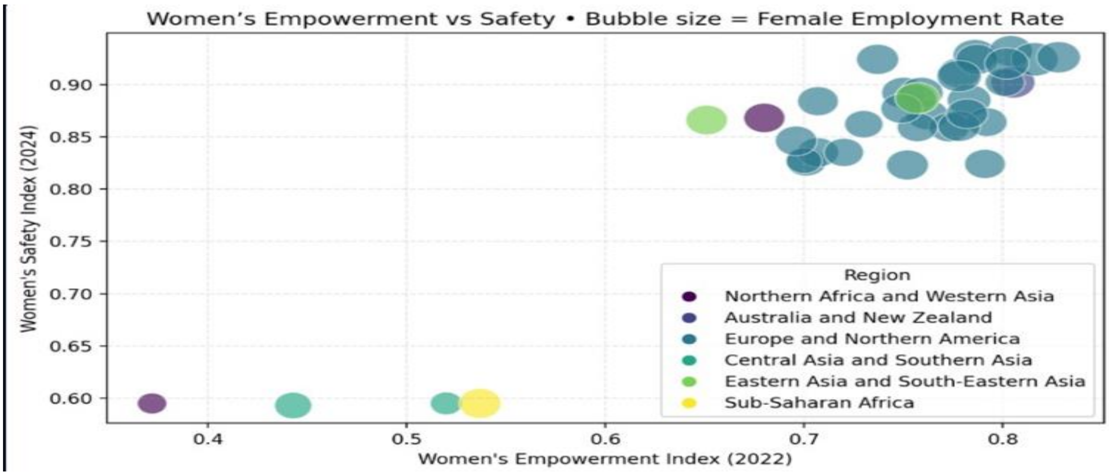


Figure 11: Safety vs Empowerment – Bubble Chart.

Insights:

- Countries with higher safety scores generally exhibit higher empowerment levels, forming a clear upward trend. This suggests that safety and empowerment often progress together.
- Larger bubbles representing higher female employment tend to appear in the upper-right region, indicating that female workforce participation is typically stronger in countries where women feel safer and more empowered.
- Nordic and Western European nations cluster tightly in the high-safety, high-empowerment area. These countries also show noticeably larger bubbles, reflecting strong employment participation.
- Regions such as South Asia, Sub-Saharan Africa, and parts of Latin America show wider variation, with many countries positioned toward mid or lower safety and empowerment levels.
- The bubble chart effectively highlights global inequality, where only a few regions consistently achieve high scores across all three dimensions.

This visualization provides a holistic view of global conditions, allowing users to identify patterns and regional clusters quickly.

6.4.2 Global Chart with India Highlighted:

In the second bubble chart, India is visually highlighted to compare its position relative to global trends.

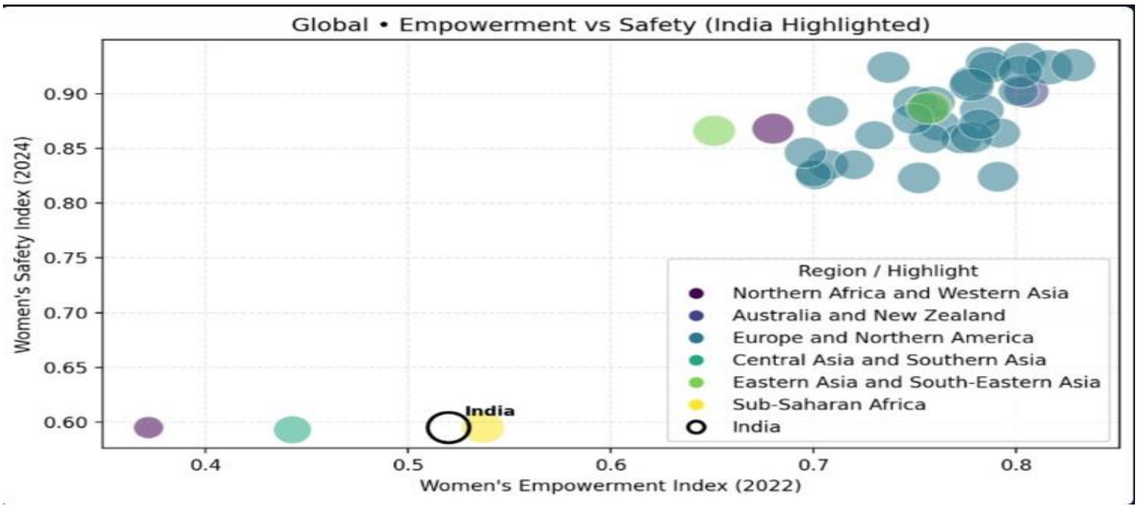


Figure 12: Safety vs Empowerment – Bubble Chart with India Highlighted.

Insights:

- India appears in the mid-range of the Women's Empowerment Index, indicating that progress has been made but remains below global leaders.
- On the Safety Index axis, India falls in the mid-to-lower range, demonstrating ongoing challenges related to women's safety an issue frequently reflected in global Gender Gap and Safety reports.
- The bubble size shows that female employment in India is moderate but lower than regions such as Europe, East Asia, and Oceania. This further reflects the gap between literacy gains and actual workforce participation.
- India aligns more closely with other developing economies, forming a cluster where empowerment is improving, but safety conditions and employment participation remain comparatively restricted.

This chart clearly illustrates that while India demonstrates progress across several women empowerment indicators, considerable gaps persist when compared with high-performing countries worldwide.

6.5 Machine Learning Prediction —Female Employment Rate (Random Forest Regression)

To extend the analytical capability of the dashboard, a Machine Learning (ML) component was integrated using a Random Forest Regression model, trained on five key indicators: *Female Employment (lag-1), Female Literacy, Female Population %, Empowerment Index (WEI 2022), and Safety Index (2024)*. This model predicts the next-year female employment rate based on the most recent socio-economic conditions of each country. Below are three sample predictions (India, Bangladesh, Italy) along with their historical trends and ML-based projections.

6.5.1 ML Prediction for India:

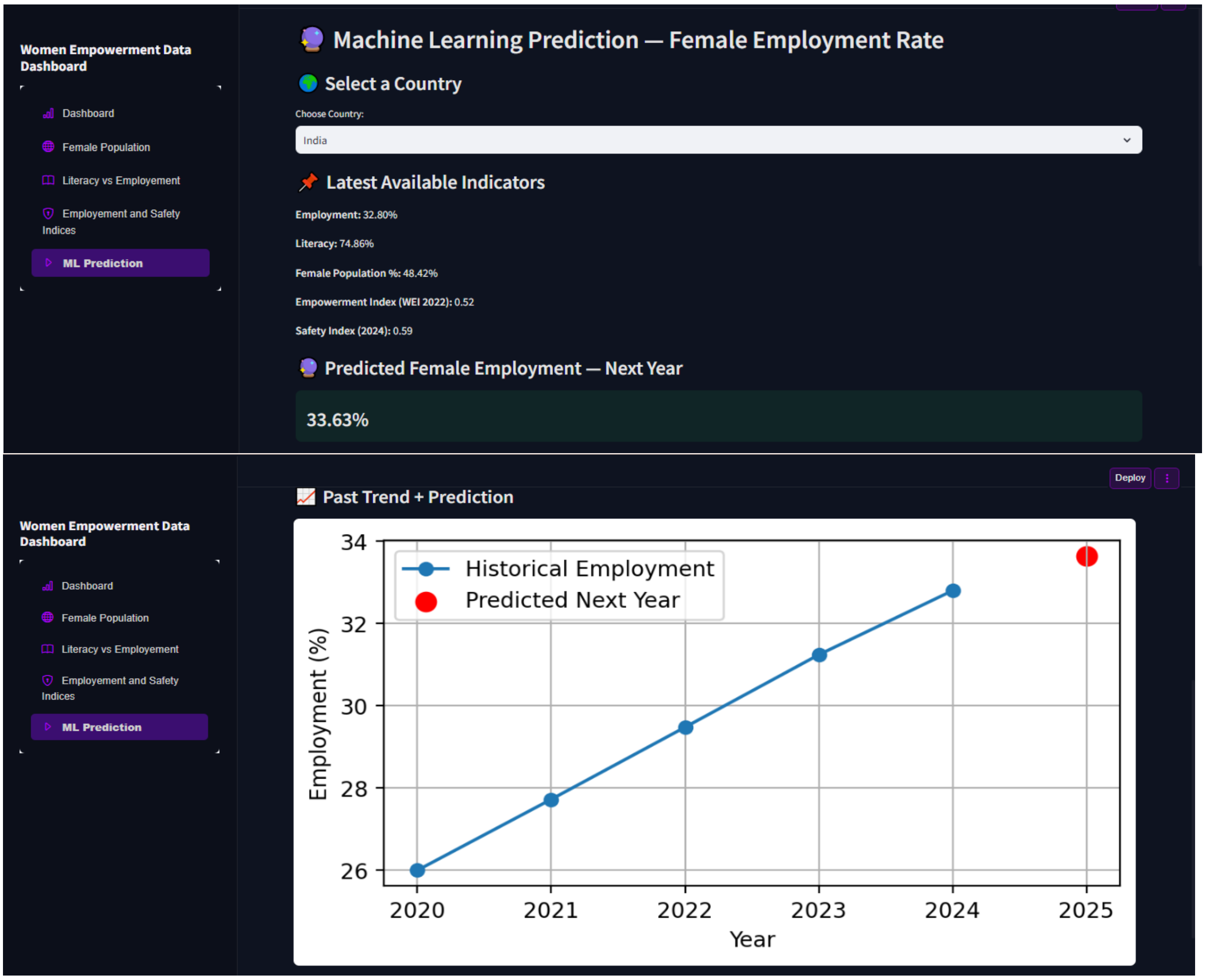


Figure 13: India — ML Predicted Female Employment Rate & Trend (2020–2025)

Insights:

- India’s female employment rate shows a steady upward trend from 2020 to 2024, rising from around 26% to 33%, indicating gradual improvement in workforce participation.

- The ML model predicts a further increase to 33.63% in 2025, which is consistent with the past trend and suggests continued positive momentum.
- The prediction is influenced by India's strong literacy progress (74.86%) and moderately improving empowerment index (0.52).
- However, India's safety index (0.59) remains a limiting factor, indicating that improvements in safety could accelerate employment growth even more.

6.5.2 ML Prediction for Bangladesh:

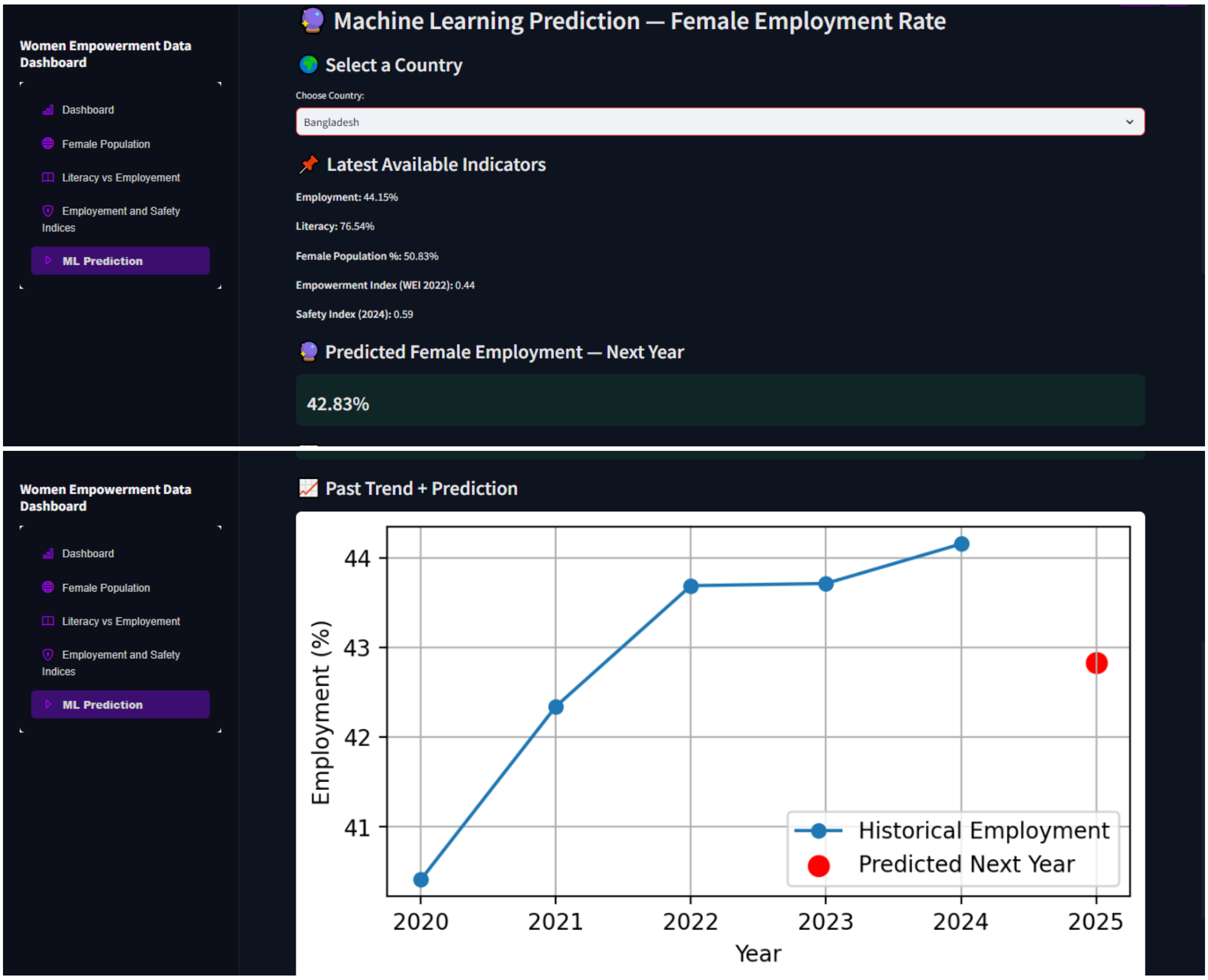


Figure 14: Bangladesh — ML Predicted Female Employment Rate & Trend (2020–2025)

Insights:

- Bangladesh shows a consistently high female employment rate, increasing from 40.7% (2020) to 44.2% (2024).
- The ML model predicts a slight decrease to 42.83% in 2025, indicating a possible stabilization or mild correction after rapid growth in previous years.
- Factors contributing to the dip include a comparatively lower empowerment index (0.44) and safety score (0.59), which may restrict further expansion.

- Despite this small decline, Bangladesh remains ahead of many South Asian countries, with one of the region’s strongest female workforce participation levels.

6.5.3 ML Prediction for Italy:

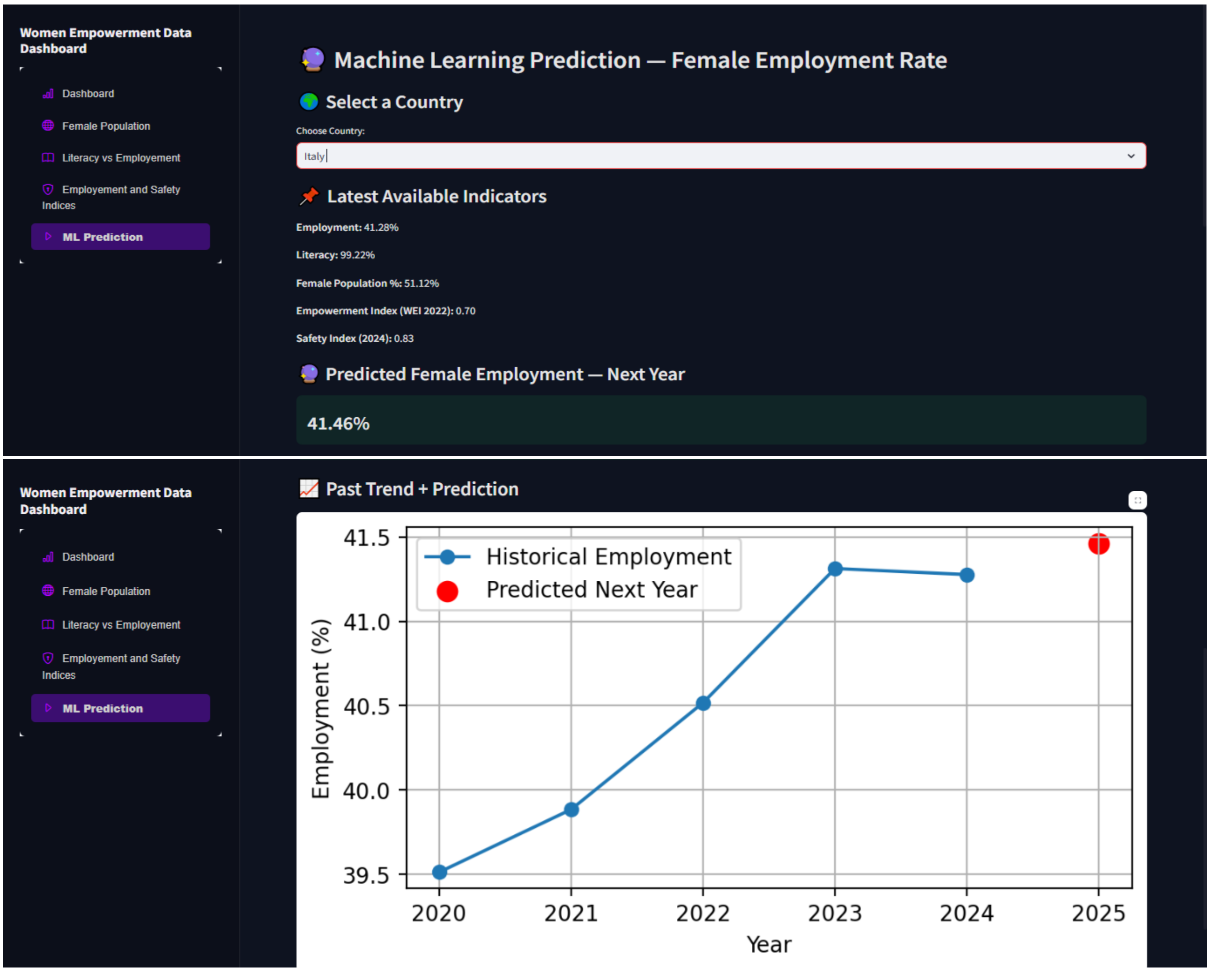


Figure 15: Italy — ML Predicted Female Employment Rate & Trend (2020–2025)

Insights:

- Italy shows a steady rise from 39.5% (2020) to 41.3% (2024) with minor year-to-year fluctuations.
- The ML model predicts an increase to 41.46% in 2025, suggesting slow but stable growth in female employment.
- Italy’s prediction is strongly supported by its high literacy rate (99%), strong empowerment index (0.70), and the highest safety index among the three countries (0.83).
- The slight upward projection indicates that Italy is maintaining a balanced socio-economic foundation that supports continued female participation in the labor market.

7. Discussion

The development of the "Global Women Empowerment Data Dashboard" helped us understand how data science and basic machine learning can be applied to study real-world social issues. By combining datasets from the World Bank, UN Women, and GIWPS/PRIO, the dashboard clearly showed differences in literacy, employment, empowerment, safety, and female population across countries. These patterns were easy to observe through interactive visualizations, which made the analysis more meaningful than simply reading raw data.

One important observation was the gap between literacy and employment in many regions, including India. Although women's literacy has increased over the years, employment levels still rise slowly, suggesting that social and economic barriers continue to affect women's workforce participation. The safety-empowerment bubble chart also showed that countries with higher safety scores usually have stronger empowerment and employment outcomes, highlighting how important security is for women's progress.

The machine learning model added an important predictive layer to the dashboard. Using Random Forest Regression, we generated next-year female employment predictions for each country. The prediction results matched the patterns seen in the visual trends for example, India showed gradual improvement, Bangladesh displayed slight fluctuations, while countries like Italy had a more stable employment pattern. This made the dashboard not only informative about the present but also useful for understanding future possibilities.

From a technical point of view, the project taught us how to clean and merge datasets with different formats, handle missing values, and connect machine learning models with a Streamlit interface. Building interactive pages and integrating ML outputs helped us learn how real dashboards are designed and how user-friendly visualizations improve data interpretation.

Overall, the discussion shows that combining visualization and machine learning can reveal hidden links among global women empowerment indicators. The project successfully demonstrated how data-driven dashboards can support research, awareness, and decision-making in a simple and accessible way.

8. Conclusion and Future work

8.1 Summary of Key Findings:

The “Global Women Empowerment Data Dashboard” helped us understand how literacy, employment, safety, empowerment, and female population collectively influence women’s development across countries. The visual comparisons clearly showed that even though literacy has improved in many regions, women’s employment still remains low, proving that education alone is not enough to ensure workforce participation. The safety–empowerment chart highlighted a strong link between security and empowerment, with Nordic countries performing the best, while South Asian and African nations showed lower values. India’s trends showed steady progress but still lagged behind global leaders.

With the addition of Machine Learning, the dashboard became more insightful by predicting next-year female employment rates using a Random Forest Regression model. These predictions matched well with recent trends for countries like India, Bangladesh, and Italy, giving users a deeper understanding of possible future outcomes. Overall, the project showed how interactive visualizations combined with ML predictions can simplify complex data and reveal important patterns in global gender inequalities.

8.2 Future Work:

While the current version of the dashboard provides strong visual, analytical, and predictive capabilities, several enhancements can further improve its depth, accuracy, and usability in the future:

1. **Adding More Indicators:** Future versions can include factors like political participation, healthcare access, digital inclusion, and wage differences to give a more complete picture of women’s empowerment.
2. **Stronger Predictive Analytics:** The current ML model predicts only the next year’s employment rate. In the future, multi-year forecasting models (like LSTM or XGBoost) can be added to predict trends for the next 5–10 years.
3. **Live Data Updates:** API integration with World Bank, UN Women, and GIWPS can automate data updates instead of uploading CSV files manually.
4. **Country-Specific Dashboards:** Detailed country pages with yearly trends, indicator breakdowns, and ML-based insights can make the dashboard more personalized and informative.
5. **Advanced Statistical Methods:** Techniques like clustering and PCA can be used to group countries with similar empowerment patterns and uncover hidden relationships.
6. **Improved User Interface:** Making the dashboard fully responsive for all devices and adding filters for region-wise and year-wise comparison can enhance user experience.
7. **Policy Insights Module:** A recommendation system can be added to provide policy suggestions based on gaps identified in the visualizations and ML predictions.

9. Reproducibility and Code availability

All datasets used in this project, including female literacy rate, employment rate, empowerment index, safety index, and population data, are publicly available from trusted sources such as the World Bank, UN Women, Kaggle, and the Women, Peace, and Security Index (WPSI). The complete source code of the dashboard covering data preprocessing, visualization modules, Machine Learning prediction scripts, and custom CSS design was developed in Python using Visual Studio Code (VS Code) as the main programming environment. Jupyter Notebook was used only for testing smaller code snippets, verifying dataset structure, and validating the ML preprocessing pipeline. To ensure reproducibility:

• All datasets are kept in CSV format so the entire workflow remains transparent and easy to reload.

• Each dashboard page and the ML model (Random Forest Regression) is handled through separate, well-structured Python functions for clarity and modularity.

• The codebase can be executed on any system that has Python, Streamlit, and scikit-learn installed, ensuring that the dashboard and the ML Prediction section can be reproduced exactly on any compatible system.

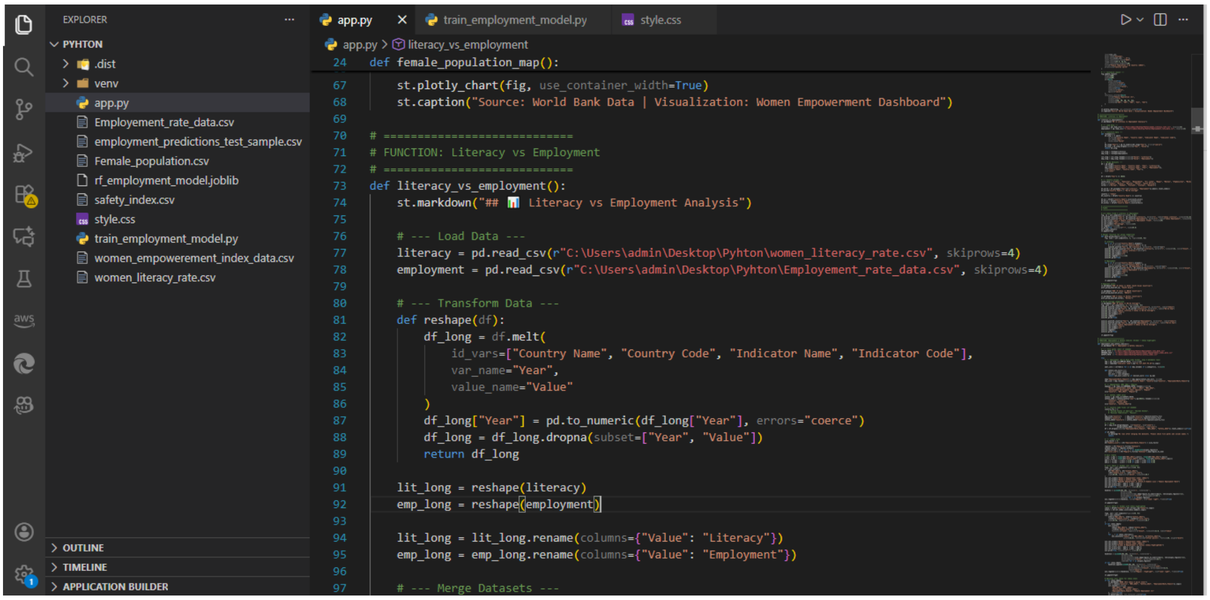


Figure 16: Code availability.

To ensure transparency and easy verification of our results, the entire project including the Streamlit dashboard, Machine Learning model, datasets, and custom styling files has been uploaded to a public GitHub repository.

Repository Link:
https://github.com/Siddiqua2007/Global-Women-Empowerment-Data-Dashboard.git

The repository contains:
• **app.py**: main Streamlit application with integrated ML prediction module
• **style.css**: complete dashboard styling
• **Female_population.csv**
• **women_literacy_rate.csv**
• **Employement_rate_data.csv**
• **women_empowerement_index_data.csv**
• **safety_index.csv**
• **rf_employment_model.joblib**: trained Random Forest Regression model used for next-year employment prediction
• **Project documentation (PDF)**

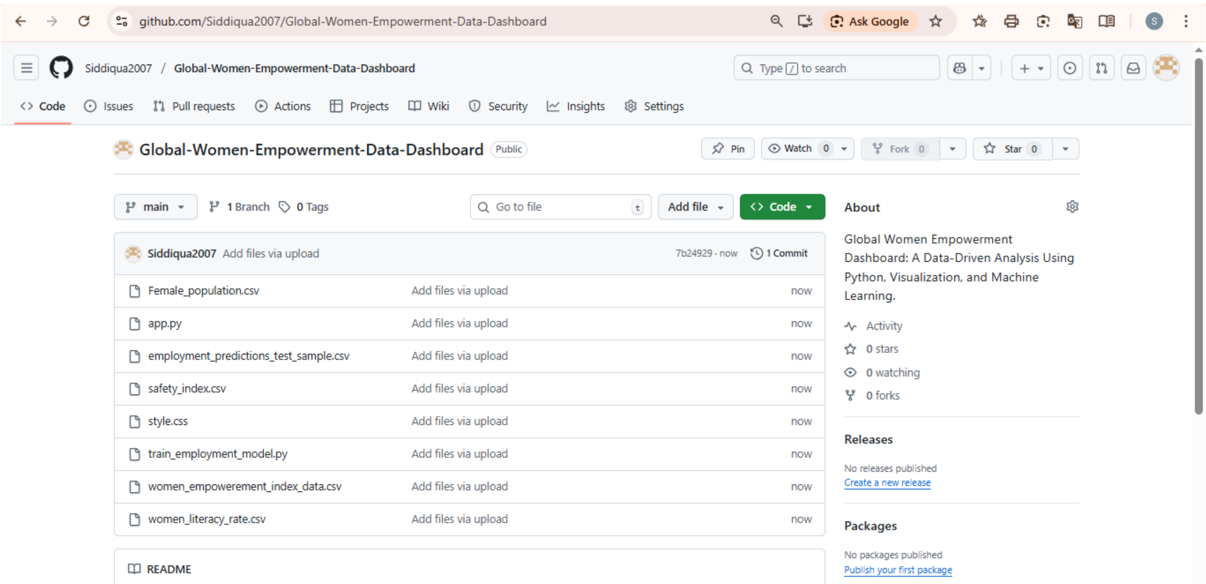


Figure 17: ML-Dashboard Repository Screenshot

This structure ensures that anyone can clone the repository, install dependencies, and reproduce the full dashboard and ML results exactly as implemented.

References

[1] A. Author, “AI and Machine Learning in Women’s Healthcare,” *Women’s Healthcare: Exploring the Potential of AI/ML*, 2023.

[2] A. N. Rana, *Leveraging Big Data to Advance Gender Equality*. IFC / World Bank Group, 2020.

[3] P. Author, “Women’s Empowerment Through AI: Predictive Safety Systems,” in *Women’s Empowerment Through Technology*, IGI Global, 2024.

[4] C. Kotero, *Machine Learning for Women’s Empowerment – Interim Report*. iMerit Technology Services, 2015.

[5] Kaggle, “Women Empowerment Indicators Dataset,” Kaggle Datasets Collection, 2024. Available: https://www.kaggle.com/

[6] World Bank, “Gender Data Portal – World Bank Gender Indicators,” The World Bank Group, 2024. Available: https://genderdata.worldbank.org/