



SDG 1 – Female Poverty Dashboard

Author: Mohamed Elshami

Programme: BSc Applied Data Science & Artificial Intelligence – Breda University of Applied Sciences (BUas)

Tools: Power BI | Python (Pandas, NumPy) | CRISP-DM Framework

Date: October 2023

1. Introduction

Despite global efforts to eradicate poverty, a significant portion of the world's population continues to grapple with extreme poverty, denying them access to basic necessities, opportunities, and a life of dignity.

The Sustainable Development Goal “No Poverty” sets an ambitious agenda to end poverty in all its forms everywhere by 2030.

However, this goal remains elusive for many, and the persistence of poverty poses a profound threat to the well-being of individuals and societies alike.

The challenge is not only to measure the extent of female poverty accurately but also to implement effective strategies that address its root causes, systemic barriers, and intersecting inequalities.

To achieve the *No Poverty* SDG, it is imperative to identify underlying factors, explore innovative solutions, and mobilize global resources to create a world where poverty is a relic of the past and where every person, regardless of gender or region, has the opportunity to thrive.

2. Research Questions

1. What are the variations in the female poverty rate across different regions?
 2. Which region has the highest poverty rate for women?
-

3. Data Sources and References

- UN Women, UNDP and Pardee Center for International Futures (using the International Futures modeling platform)
 - World Bank – Poverty and Inequality Platform (PIP)
-

4. Data Understanding and Acquisition

Which data exists in the world?

The amount of data that exists globally is vast and growing. It includes structured data (databases, spreadsheets), unstructured data (text, images, videos), big data (large complex datasets), open data (publicly available datasets), and personal data (individual records).

Which data is needed?

For this project, data related to **female poverty rates across different regions** was required.

This includes indicators such as the share of the female population living below various poverty lines, income levels, and other socioeconomic factors.

How can the data be acquired?

Data was acquired from **publicly available, reputable sources**, primarily the **World Bank** and **UN Women/UNDP** databases. These contain global poverty statistics collected through household surveys and international reports.

5. Data Cleaning and Preparation

Task	Action & Reason
Use first row as header	Ensures correct column names for further processing.
Rename tables/columns	Simplified <i>[Figure 3a and 3b]</i> → <i>female poverty rate</i> for clarity.
Convert data types	\$1.90, \$3.20, \$5.50 from text to percentage for correct calculations.
Delete unnecessary columns	Removed irrelevant columns to improve efficiency.
Handle missing values	Removed rows with NaN values to avoid bias.
Remove duplicates	Ensured data integrity and accuracy.
Scale data	Scaled \$1.90/\$3.20/\$5.50 columns for comparability across regions.

6. Exploratory Data Analysis (EDA)

Distribution Analysis

The data distribution revealed whether values were symmetric or skewed. Right-skewed patterns indicated many regions with low female poverty and a few with extremely high rates.

Summary Statistics

- *Mean* – Average female poverty rate.
- *Median* – Central value, less affected by outliers.
- *Standard Deviation* – Variation between regions.

- *Min & Max* – Overall range of poverty rates.
- *Quartiles* – Spread and outlier identification.

Visualization Techniques Used

- *Histograms* – Distribution overview.
 - *Box Plots* – Spread and outliers (used on Female Extreme Poverty Headcounts page).
 - *Scatter Plots* – Correlations between thresholds.
 - *Maps* – Geographical distribution of female poverty worldwide.
-

7. Dashboard Visualizations and Findings

Bar Chart Visualization

Title: Female Poverty Rate by Region

Description: Displays percentages of women living below \$1.90, \$3.20, and \$5.50 per day thresholds.

Key Observations:

- Sub-Saharan Africa has the highest female poverty rates at all thresholds.
- Oceania has the second highest rates.
- Europe and Northern America show the lowest levels.

Interpretation:

Large disparities highlight regional economic inequalities and differences in poverty alleviation programs.

Targeted interventions are needed for regions with severe female poverty.

Box Plot Visualization

Title: Distribution of Female Poverty Rate Living on \$1.90 Per Day Across Regions

Description: Shows median and variability of female poverty rates by region.

Key Observations:

- Half of regions have female poverty rates above 20%.
- Some regions exceed 40%, revealing extreme inequality.

Interpretation:

Female poverty is widespread and varies sharply by region, indicating critical social and economic disparities.

Correlation Analysis Visualization

Title: Correlation Between Female Poverty Rates at Different Thresholds

Description: Correlation coefficient ≈ 0.94 between \$1.90 and \$3.20 rates.

Key Observations: Regions with high rates at \$1.90 also tend to be high at \$3.20.

Interpretation: Strong positive correlation suggests consistency in female poverty across thresholds; interventions at lower poverty levels will also aid those slightly above them.

8. Discussion

In my analysis of female poverty rates across regions, I identified several key considerations and limitations.

- **Missing Data:** Some economic indicators were absent, risking bias and reducing representativeness.
- **Skewed Distribution:** Uneven data complicates normalization and statistical tests.
- **Collection Bias:** Underreporting may exist where women lack access to surveys.
- **Regional Differences:** Government policies and cultural factors affect comparability.

To mitigate these issues, I propose imputation for missing values, sensitivity analysis for bias testing, and re-evaluation of skewed data through transformation techniques.

Understanding these limitations ensures a more balanced interpretation of female poverty worldwide and supports data-driven decision-making.

9. Recommendations for Future Analyses

To improve robustness and depth:

- Use correlational and linear regression analysis to explore causal factors.
- Follow the CRISP-DM framework for iterative refinement:
 1. **Business Understanding** – Refine research focus.
 2. **Data Understanding** – Gather additional regional indicators.
 3. **Data Preparation** – Enhance dataset quality and consistency.

4. **Modeling** – Apply advanced techniques for trend analysis.

5. **Evaluation & Deployment** – Present insights for policy use.

This structured approach will enable continuous improvement and stronger alignment with SDG 1 objectives.

10. Conclusion

Summary of the Problem and Research Question

The study focused on female poverty and regional disparities.

Core questions: How do female poverty rates vary by region, and which region shows the highest rates?

Findings of the EDA

- **Regional Variation:** Sub-Saharan Africa has the highest rates; Europe and North America the lowest.
- **Income Disparities:** Economic conditions vary drastically across regions.
- **Government Policies:** Strong social safety nets reduce poverty in developed regions.

Interpretation

Regional economic health and policy effectiveness drive female poverty outcomes.

Targeted interventions in regions with weak social protection are essential.

Limitations and Next Steps

- Address missing data and collection bias.
- Apply normalization to handle skew.
- Advance to predictive analysis and policy simulation in future CRISP-DM cycles.

Final Thoughts

This dashboard demonstrates how data-driven analysis can support the global goal of eradicating poverty.

By visualizing regional disparities in female poverty, the project provides a foundation for targeted actions and evidence-based decision making.

Continuous iterations of the CRISP-DM process will ensure that insights remain relevant, accurate, and impactful in advancing SDG 1.