USA

WOMEN IN WORKFORCE

VEDUDE RABİA SÖZKESEN 32182

DATA

USA

Share of women in the workforce

Average years of schooling for women

Total fertility rate: births per woman

Marriage rates

Average age of mothers at childbirth by birth order

HYPOTHESIS

Null Hypothesis(H0): There is no correlation between any of the 4 datasets and women in workforce.

Alternative Hypothesis(H1): There is a *pozitive correlation* between women in workforce and;

- Average years of schooling for women
- Average age of mothers at childbirth by birth order

There is a *negative correlation* between women in workforce and;

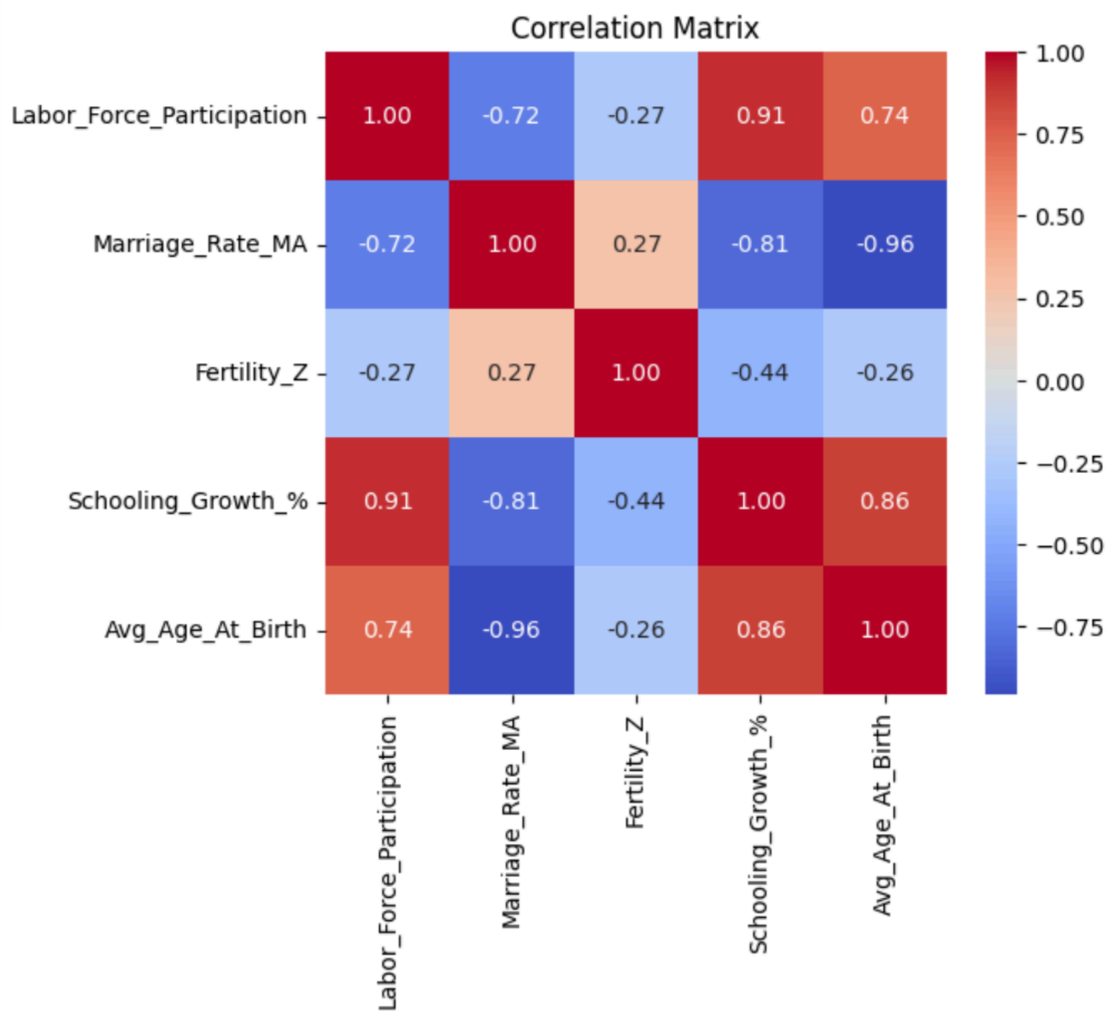
- Total fertility rate: births per woman
- Marriage rates

MOTIVATION

I want to investigate the factors that influence—and are influenced by—women’s participation in the workforce. I believe that understanding these relationships helps us better comprehend today’s world.

Identifying such trends can also guide the development of more strategic approaches to increasing women’s participation in the labor force and ensuring equal access to employment opportunities for all women.

CORRELATION MATRIX



Positive Correlations:

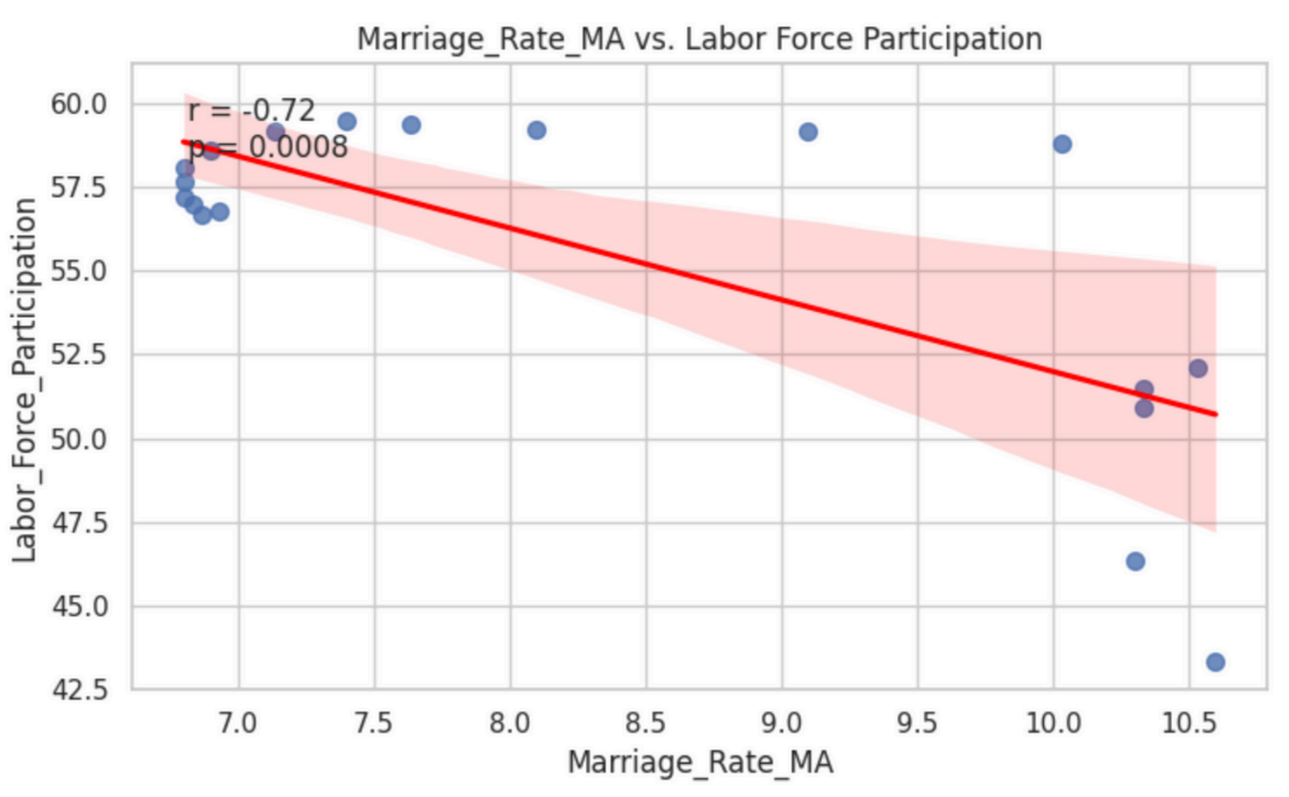
- schooling - labor force (strong)
- schooling - average age at birth (strong)
- average age at birth - labor force (weaker)

Negative Correlations:

- marriage rate - average age at birth (strong)
- schooling - marriage rate (strong)
- marriage rate - labor force (weaker)

CORRELATION GRAPHS

MARRIAGE AND WORKFORCE

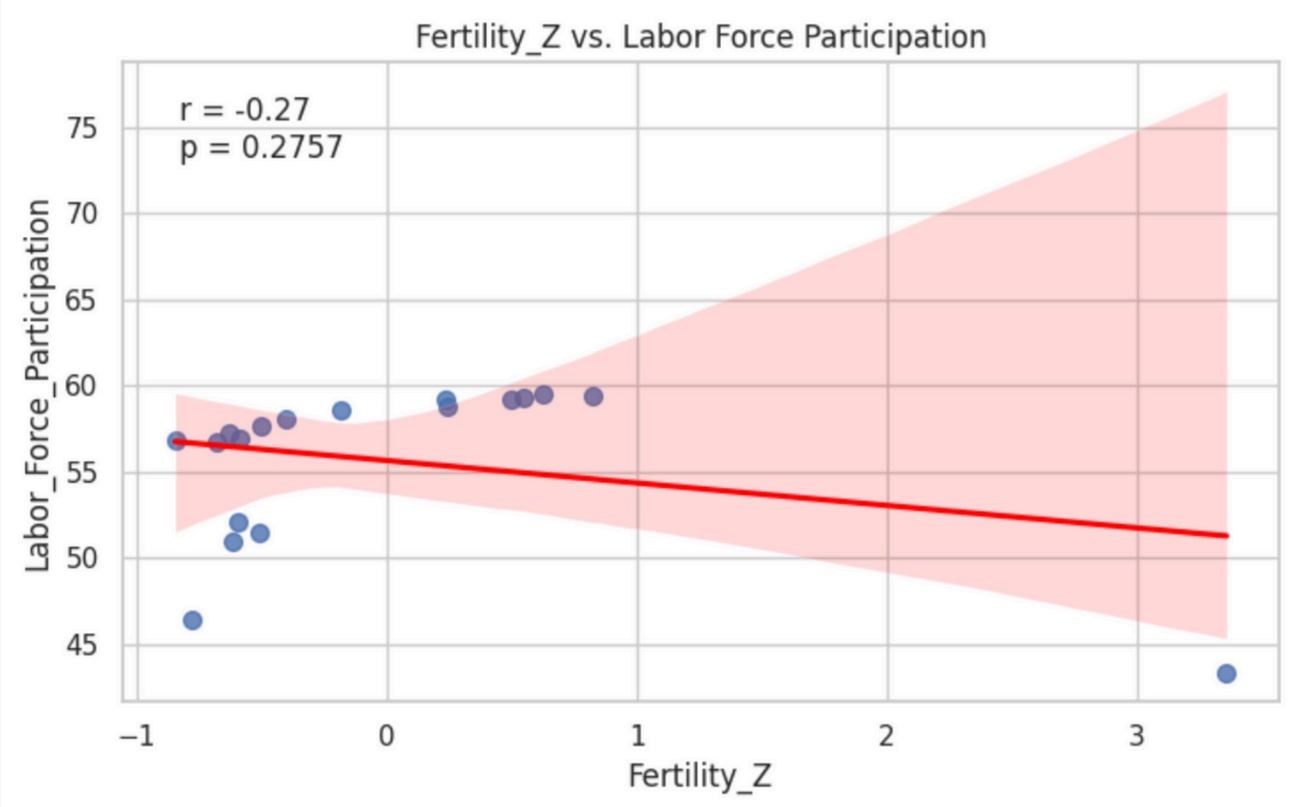


Values:
r = -0.72, p = 0.0008 (significant)

Interpretation:
There is a moderate negative correlation between marriage rates and the number of women in the workforce, suggesting that women who work may be less likely to get married.

CORRELATION GRAPHS

FERTILITY AND WORKFORCE



Values:
r = -0.27, p = 0.2757 (not significant)

Interpretation:
There is a very weak and negligible negative correlation, suggesting that working women's decision to have children may not be strongly influenced by their employment status.

CORRELATION GRAPHS

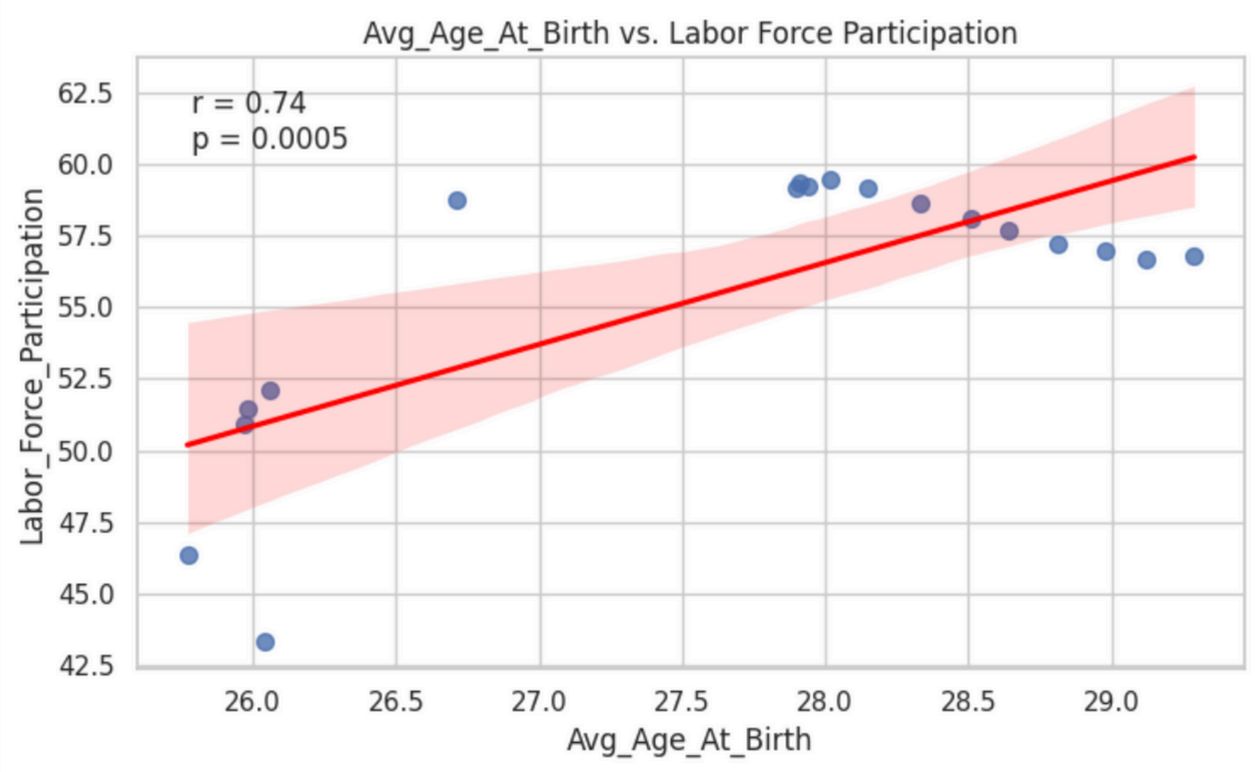
SCHOOLING AND WORKFORCE

Values:
r = 0.91, p = 0.0000 (significant)

Interpretation:
There is a strong positive correlation between women's education levels and their participation in the labor force, which may indicate that educated women are more likely to be employed.

CORRELATION GRAPHS

AGE AT BIRTH AND WORKFORCE

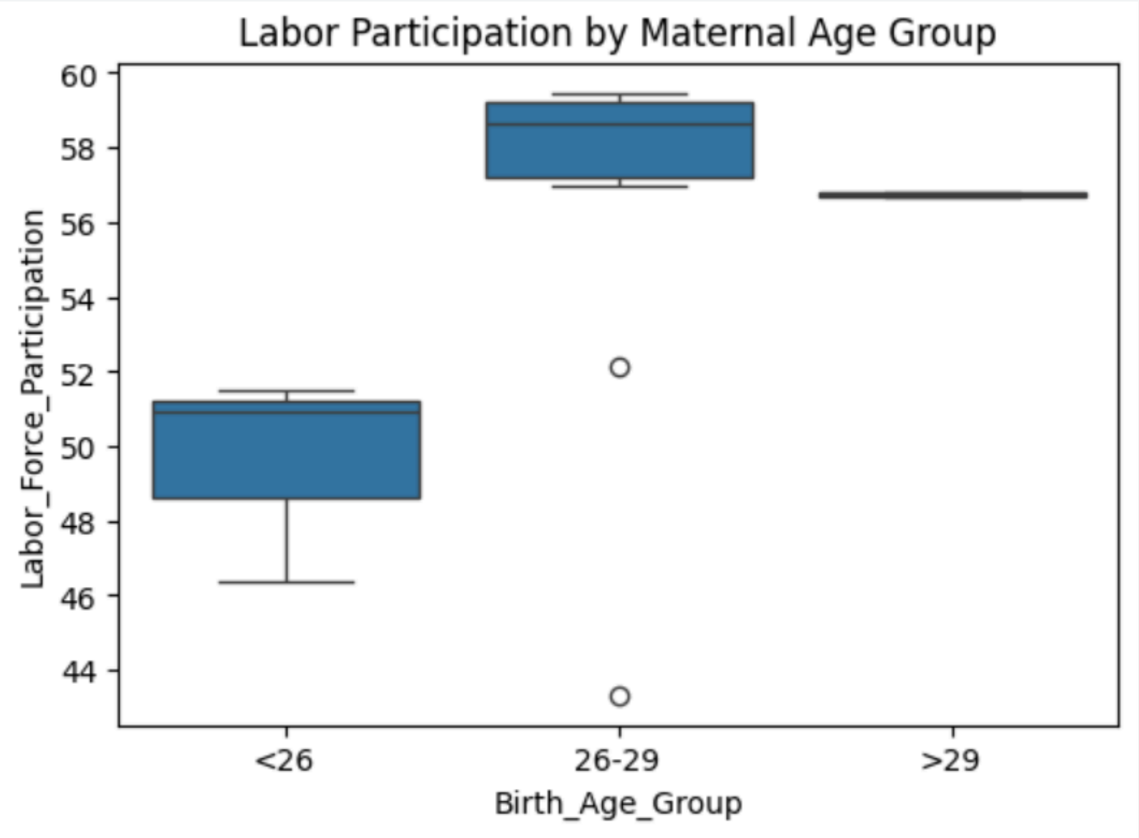


Values:
r = 0.74, p = 0.0005 (significant)

Interpretation:
There is a moderate positive correlation between the average age of mothers at childbirth and the number of women in the workforce, which may indicate that as women participate more in the labor force, they tend to have children at older ages.

CORRELATION GRAPHS

AGE AT BIRTH AND WORKFORCE

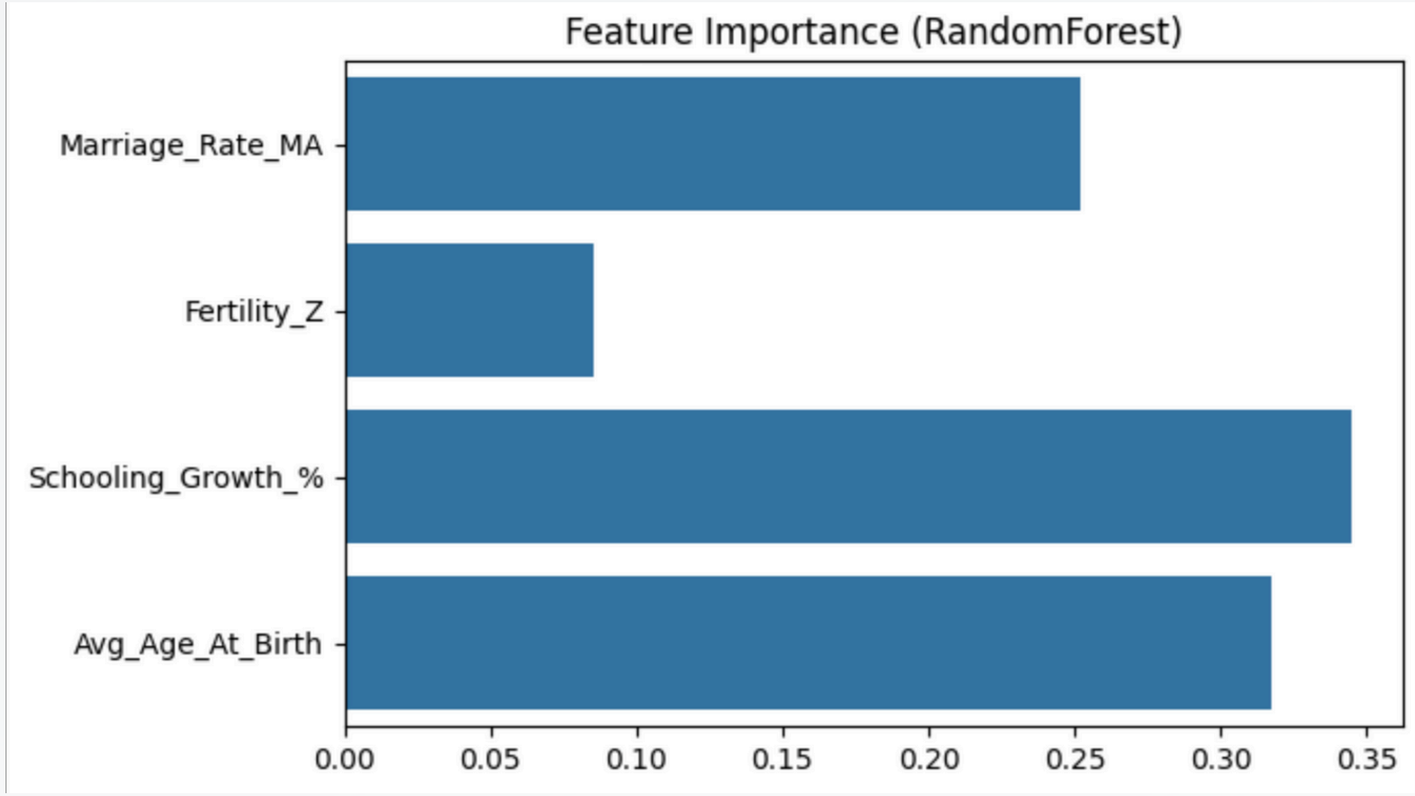


Interpretation:

This boxplot shows that women who give birth at *older* ages tend to have higher and more consistent labor-force participation. Labor participation is lowest for those *under 26*, and highest for those giving birth between *26–29 or after 29*. This suggests a positive relationship between delayed childbirth and workforce involvement.

RANDOM FOREST REGRESSION ANALYSIS

The Random Forest Regressor identified **Schooling Growth** (%) as the most important predictor of female labor-force participation (0.34), followed by **Average Age at Birth** (0.31) and **Marriage Rate** (3-year Moving Avarage) (0.25).



Fertility Rate (z-score) had the lowest importance (0.09), suggesting a minor influence. These results highlight that educational gains and delayed childbirth are strongly linked to higher workforce participation, while fertility changes have less impact.

ANALYSIS

EDA AND HYPOTHESIS TESTING

Data Transformation

- Applied a 3-year moving average on marriage rate to reduce short-term noise.
- Standardized fertility rate for scale-independence.
- Calculated percent growth in women's schooling over time to show progress.
- Binned maternal age into three meaningful age groups for easier group-level analysis.

Exploratory Data Analysis

- Visualized relationships using scatter plots with regression lines.
- Used seaborn for clean, readable plots.
- pearson correlation analysis.
- Correlation heatmap to observe variable interactions.

Hypothesis Testing

- Pearson's r and p-value to assess linear relationships and significance
- ANOVA for categorical group comparisons.

Data Cleaning & Standardization

- Renamed long or inconsistent column names for clarity.
- Located the correct schooling column.

ANALYSIS

ML METHODS

Random Forest Regressor

- An ensemble-based tree model that builds multiple decision trees and averages their outputs.
- Tuned with GridSearchCV on parameters:
  - n_estimators (number of trees)
  - max_depth (maximum depth of each tree)

Support Vector Regression

- A kernel-based regression method for capturing both linear and non-linear relationships.
- Tuned with GridSearchCV on parameters:
  - kernel (rbf or linear)
  - C (regularization strength)
  - epsilon (margin of tolerance)

FUTURE IMPROVEMENTS

Since the data spans many years, the correlations can be analyzed in relation to historical events that occurred during those periods.

The analysis can be expanded beyond the U.S. to include other regions, allowing comparisons across countries. Interpreting the data by both time and location can reveal how trends vary across different contexts.