Breaking Barriers:

Unveiling the Future of Women in STEM

Our Team:

SARA

SHARON

GRACIE

Agenda:

01 Introduction

02 Statistics

03 Our Questions and models

04 Conclusion

514,323

Men graduate with a STEM degree in the USA

276,429

Women graduate with a STEM degree in the USA

1/3

are hired after graduation

23%

higher drop out rate compared to
male counter part

Problem Statement & Methodology

Gender disparities persist in STEM fields globally, with women underrepresented compared to men.

We want to provide insights for various stakeholders into the factors that influence women's achievement in STEM fields. This is crucial for to achieving gender equality and ensuring equal opportunities in high-demand career fields.

Based on personal experience, we recognize that **equal access is not just a matter of social justice but also an economic imperative**.

DATA COLLECTION

World Bank graduates by field and gender data

PREPROCESSED DATA

Feature refinement
RobustScaler()
Initial Visualizations

DEFINED QUESTIONS

Questions based on various factors that change trajectory of women in STEM

RAN MODELS

Tested multiple models for each question and compared results (Prediction models and clustering)

ANALYZED RESULTS

Feature importances
Predict future STEM enrollment
Compare countries assigned to each cluster

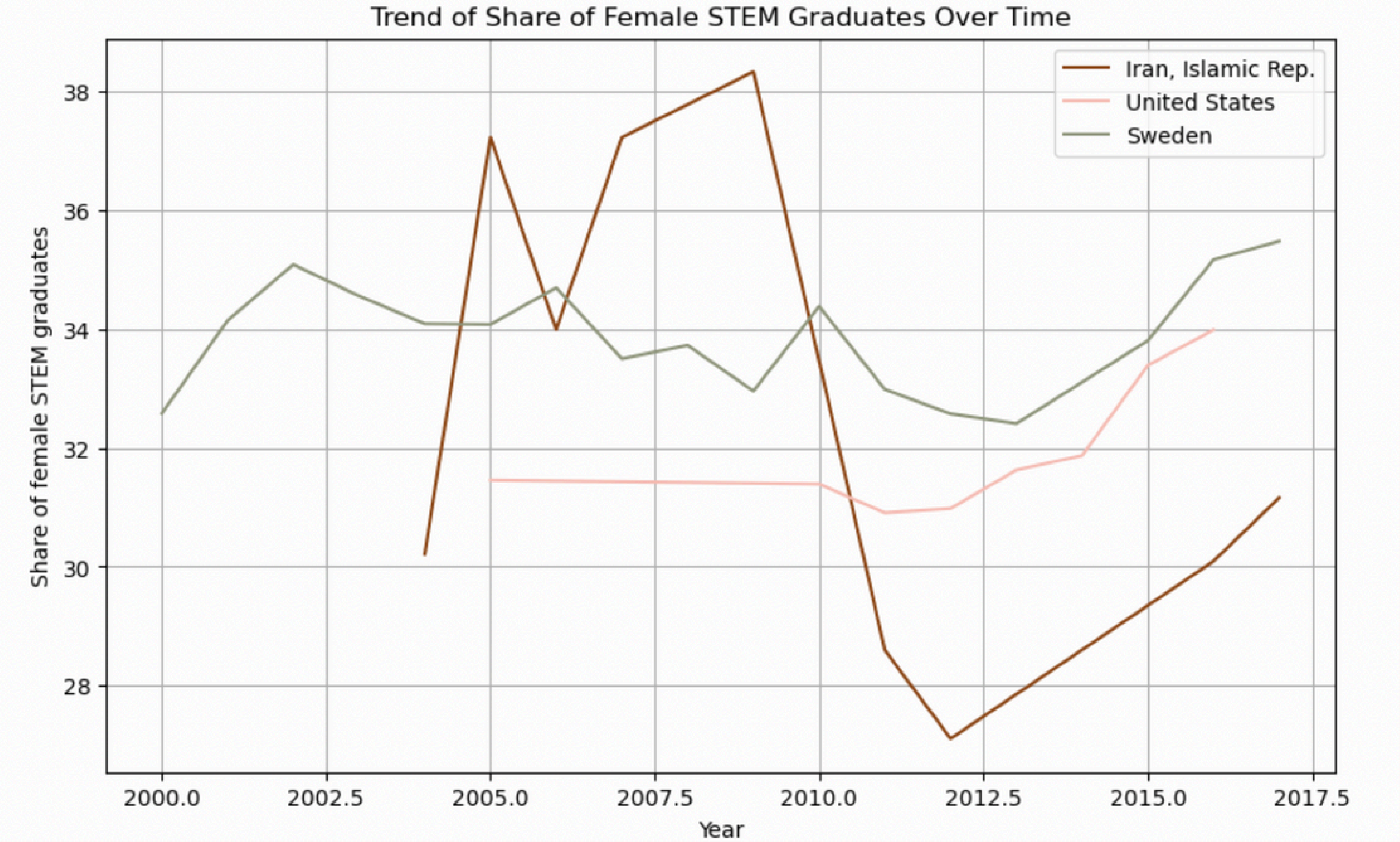
01

What is the relationship between various socio-economic factors and the share of female STEM graduates over time?

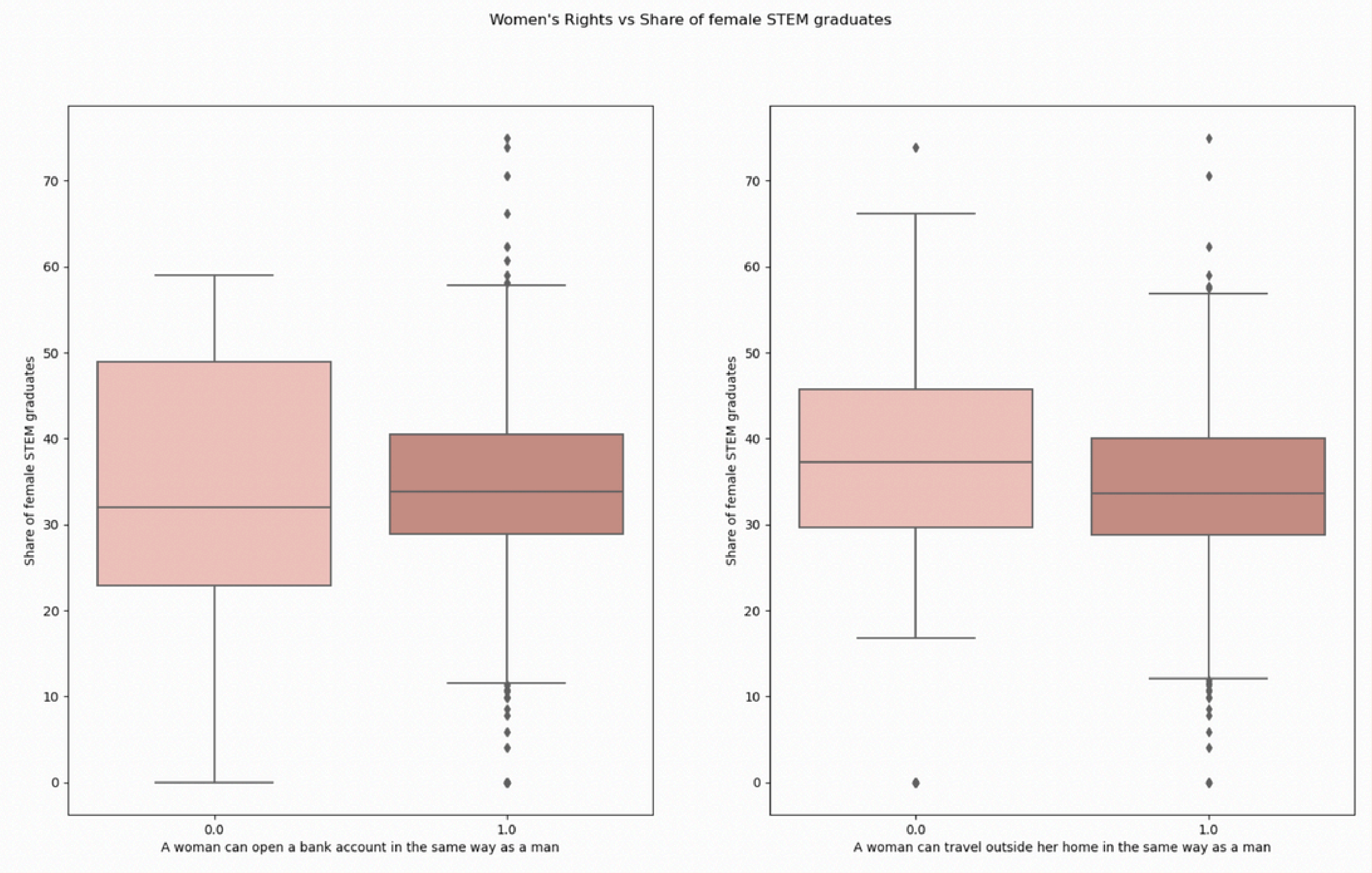
Elaborate on what you want to discuss.

Female STEM grads over time in various countries

TRENDS OF FEMALE GRADS OVER TIME



WOMENS RIGHTS CORRELATED WITH FEMALE STEM GRADS



Model of choice

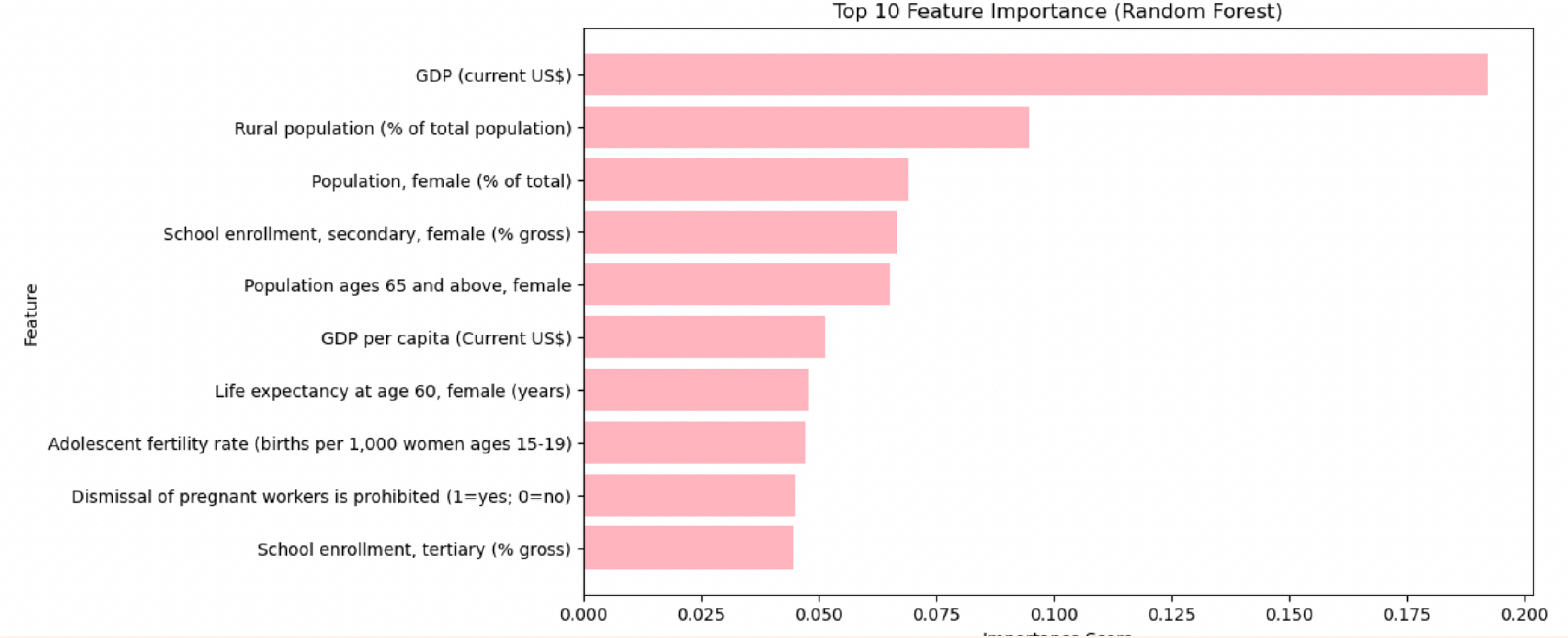
RANDOM FOREST

	Model	R Squared Train	R Squared Test	MSE Train	MSE Test
0	Linear Regression	0.362848	0.203505	0.496795	0.553359
1	Ridge	0.362228	0.214036	0.753791	0.726642
2	Ridge (Hypertuned)	0.033244	-0.045916	0.753791	0.726642
3	Random Forest	0.956012	0.638958	0.034298	0.250831
4	Random Forest (Hypertuned)	0.170447	-0.020719	0.646813	0.709138
5	ElasticNet	0.007791	0.001955	0.779712	0.694834
6	ElasticNet (Hypertuned)	0.000000	-0.000131	0.779712	0.694834

REASON

THIS MODEL HAS THE HIGHEST R SQUARED AND THE LOWEST MSE

Female Percent of Secondary School Enrollment Feature Importance



02

What factors indicate whether or not a women completes secondary education?

Model of choice

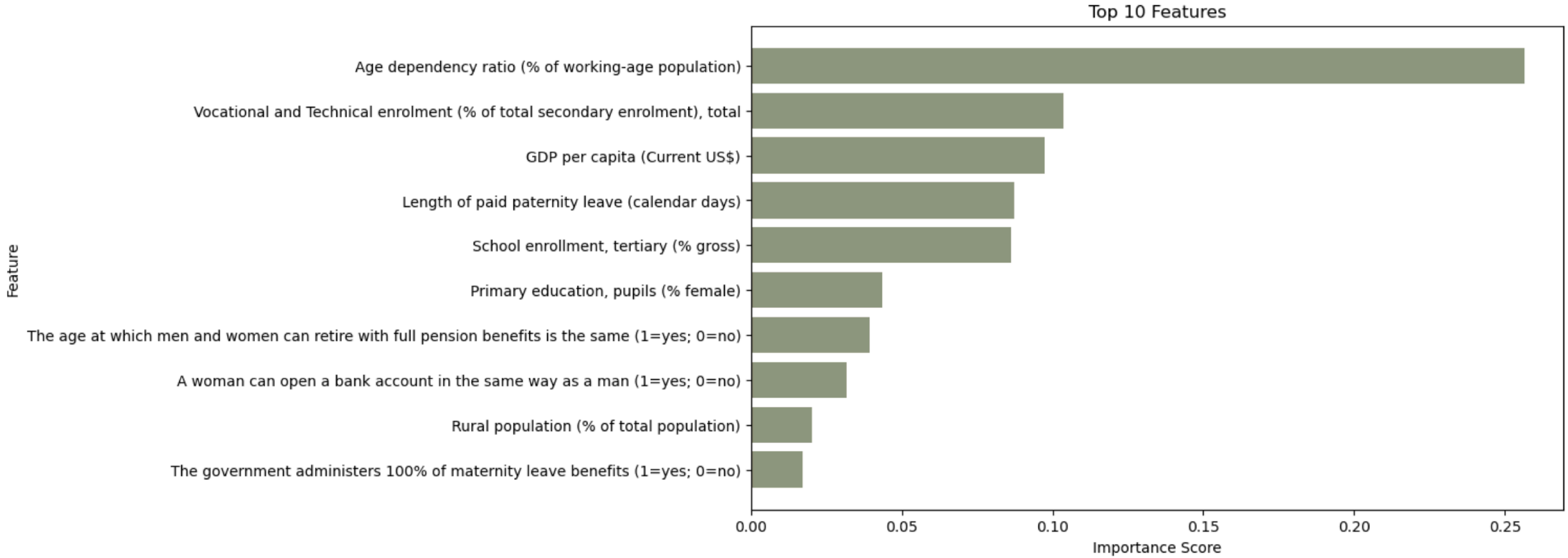
XGBOOST REGRESSOR HYPERTUNED

	Model	R Squared Train	R Squared Test	MSE Train	MSE Test
0	Linear Regression	0.788097	0.564605	0.563901	0.723251
1	Ridge (Hypertuned)	0.662649	0.578002	0.668194	0.763849
2	Random Forest (Hypertuned)	0.985725	0.802023	0.029087	0.358353
3	XGBoost (Hypertuned)	0.998486	0.830043	0.003085	0.307634

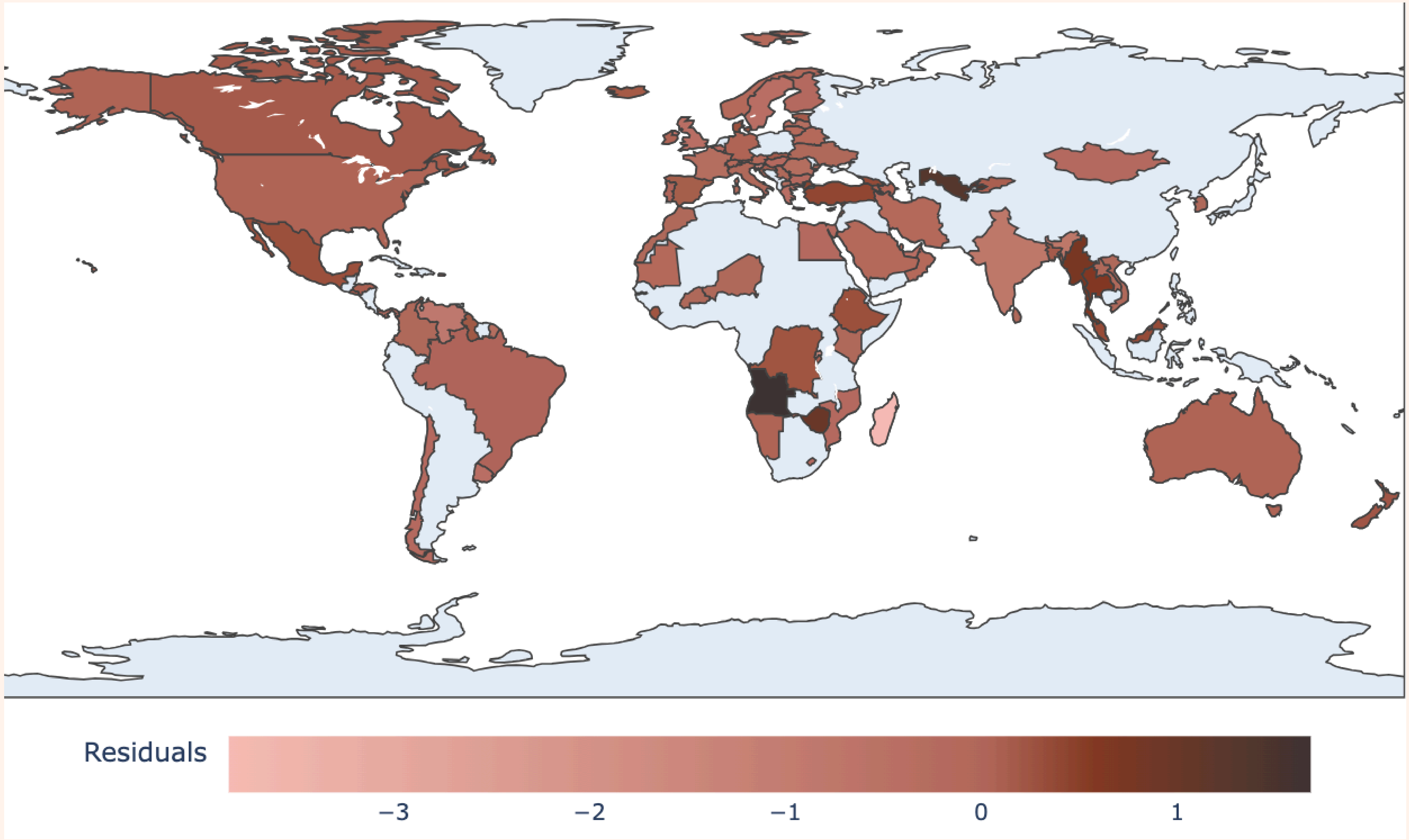
REASON

THIS MODEL HAS THE HIGHEST R-SQUARED AND LOWEST MSE COMPARED TO OTHER MODELS

Female Percent of Secondary School Enrollment Feature Importance



Residual Chloropleth Map



DIFFERENCE IN ACTUAL VALUES AND PREDICTIONS

The map to the left shows the countries that were included in the prediction models and the value of their residuals.

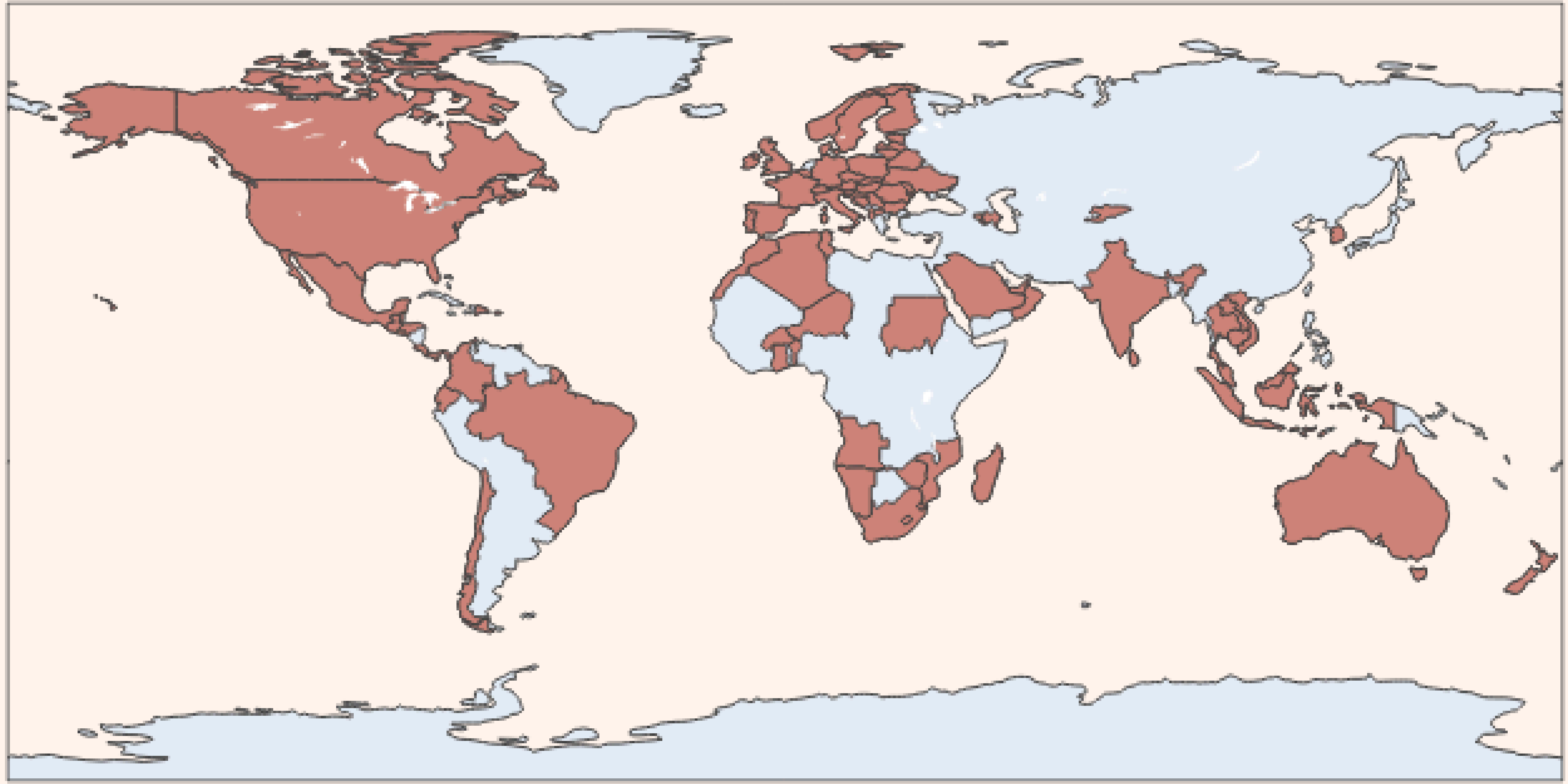
Large residuals:

- Madagascar
- Angola
- Uzbekistan
- Zimbabwe

Are there groups of countries that share characteristics leading to differences in females' graduation rates from STEM fields?

Unsupervised Learning

To explore countries with similar macroeconomic indicators, education rates among women, and/or women's rights



2015 Data Available

2015 Data Not Available

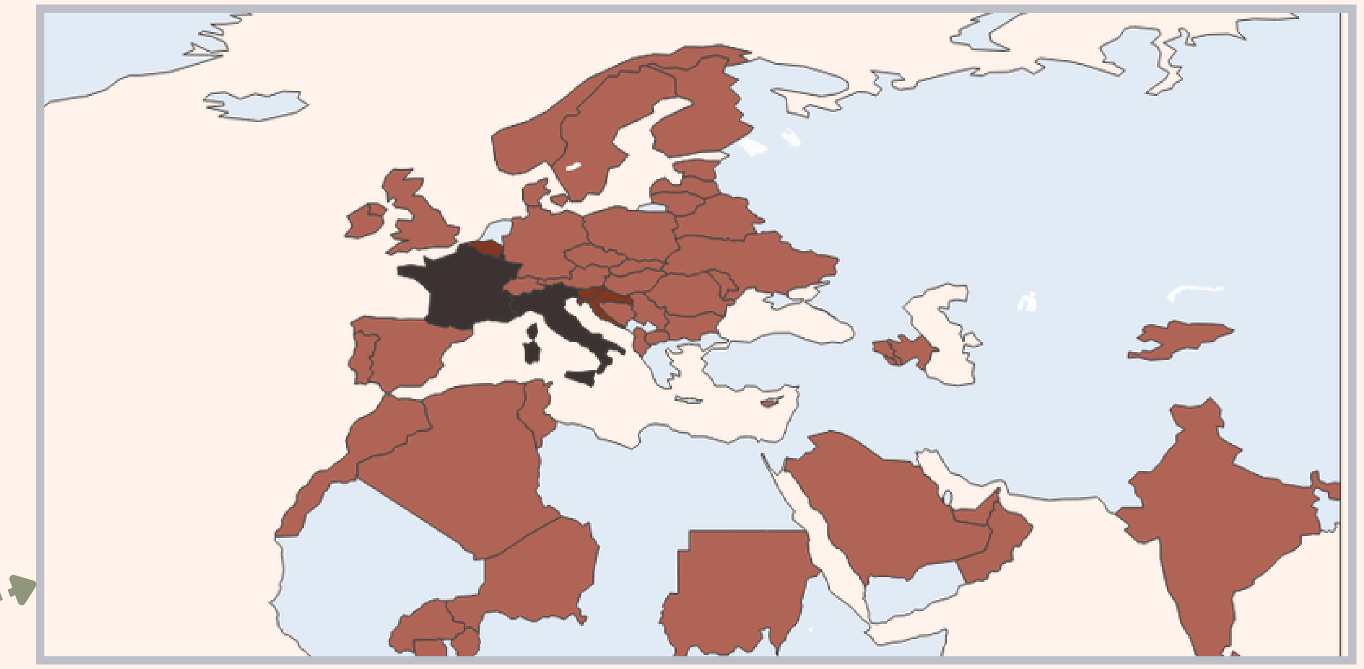
2015 DATA AVAILABILITY BY COUNTRY

To gain the most insight from clustering, data was limited to a single year, 2015.

The map to the left shows the countries that were included in the clustering models.

RESULTS (DBSCAN WITH HYPERTUNING)

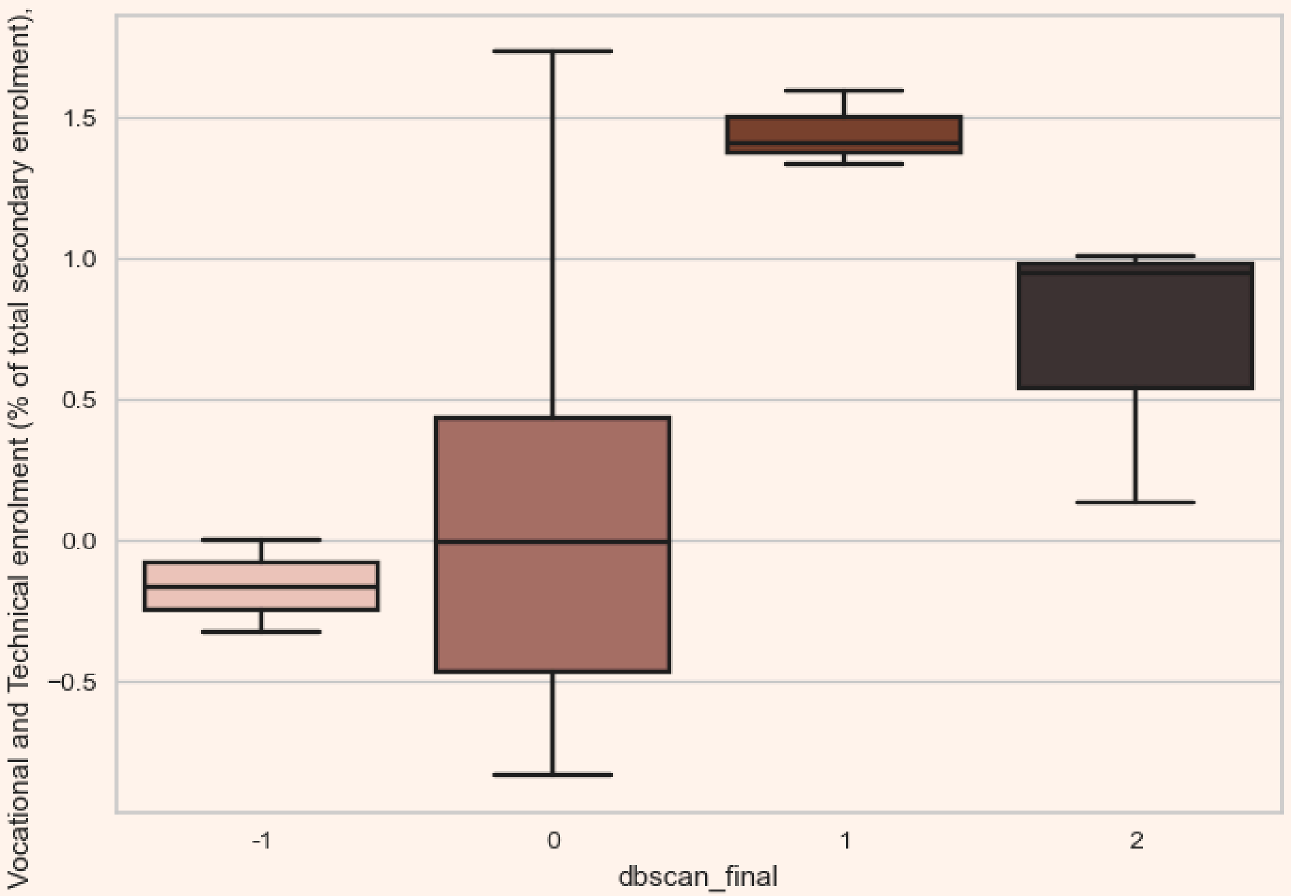
Overall, there were not obvious clusters in the data. The vast majority of countries were assigned to cluster 0, and only a handful of countries were assigned to clusters 1, 2, and 3.

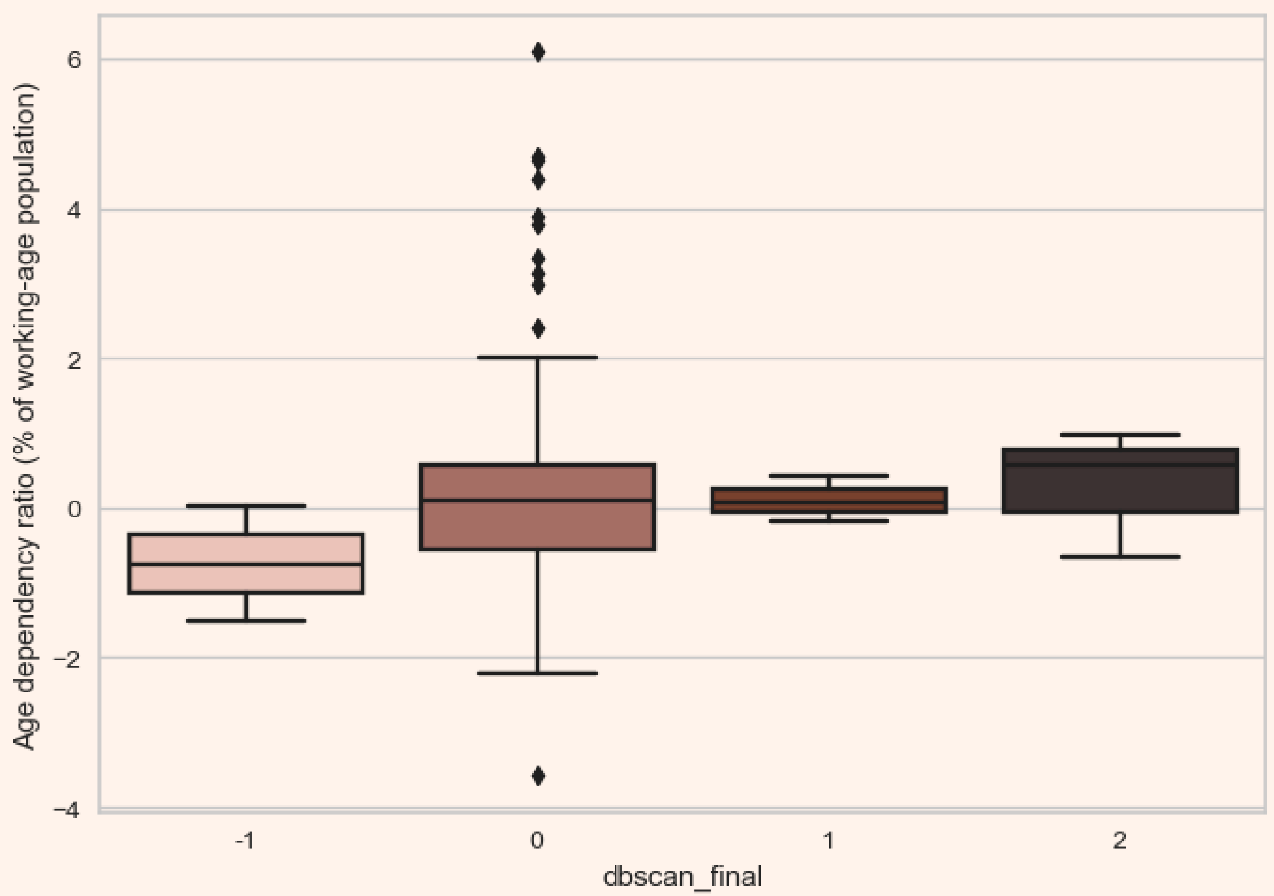


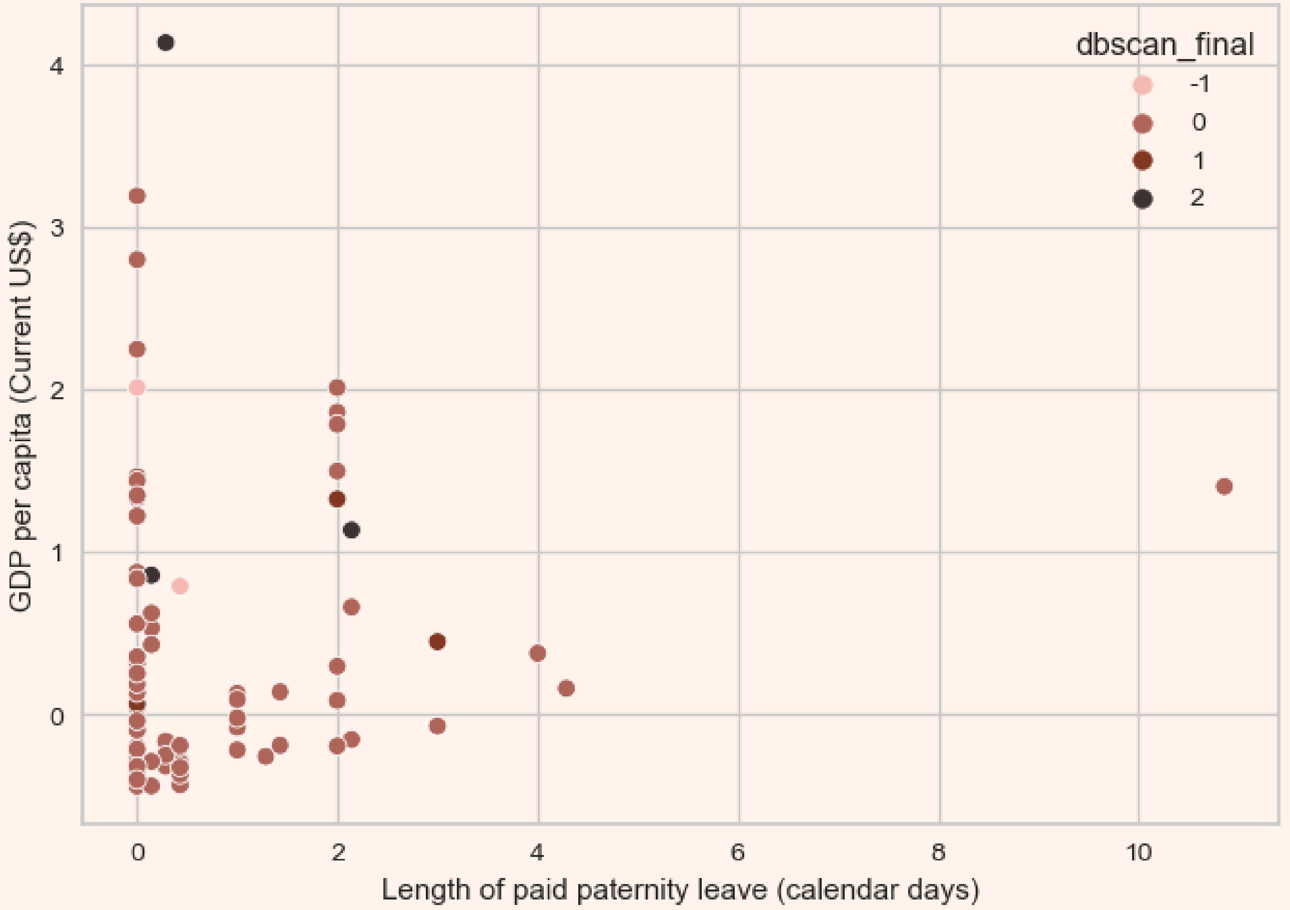
LACK OF CLEAR COUNTRY GROUPS

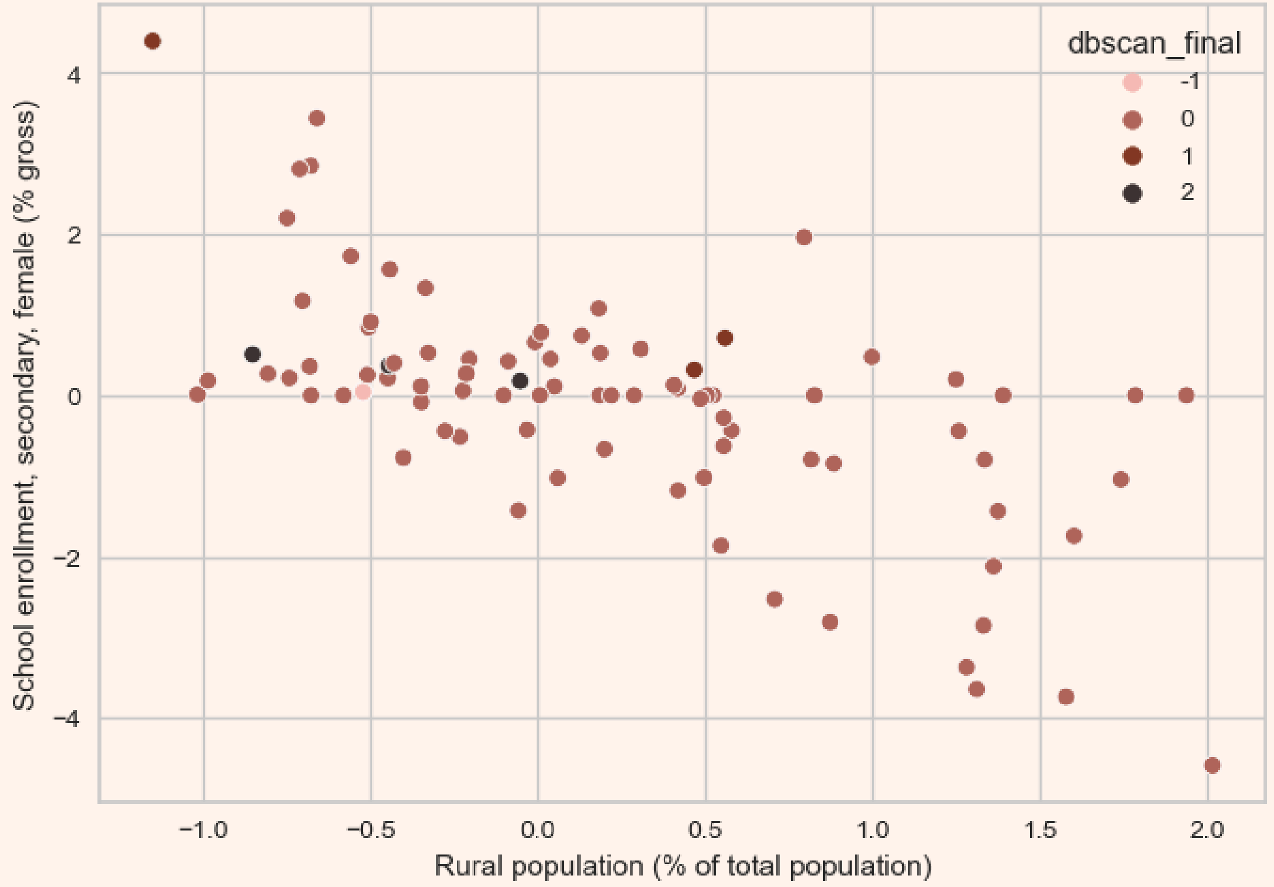
Among the most important features from previous models, the clusters seemed most clearly stratified by Vocational and Technical enrollment and Adolescent Fertility Rate.

Country Name	Cluster (DBSCAN)
Korea, Rep., United States	-1
60+ remaining countries...	0
Belgium, Croatia, Slovenia	1
France, Italy, Luxembourg	2









Future Work

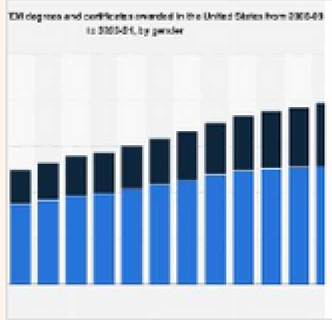
- Search for more recent data (our dataset stopped at 2020)
- Talk to Education experts to find new features
- Handle differences in data between countries
- Spend more time picking models and hypertuning

Conclusion

Women's historical underrepresentation in STEM fields globally has prompted the use of machine learning to reveal key factors impacting their pursuit of STEM education. This approach helps address biases and barriers, fostering greater gender equality and diversity in STEM disciplines.

Thank you!

Resource Page



STEM degrees awarded in the U.S. by gender 2021

In the school year 2020-21, 514,323 males were awarded STEM certifications in the United States.

Statista

- Data is in the google drive: World Data
- All other references are in our individual notebooks