CSCI 5502 - Data Mining

Mid Point Check-in

Analysis of Gender Bias in the Employment Sector

University of Colorado Boulder

Team

Amatullah Sethjiwala
amse5199@colorado.edu

Srihaasa
srpi9270@colorado.edu

Preethi
prvi9376@colorado.edu

Xuefei
xusu2102@colorado.edu

Table of Contents

Introduction

- Gender bias is greatly widespread in the education and employment sectors.
- Gender inequality incurs a loss of up to USD 12 trillion or 16% of global income
- Our work aims to do an in-depth analysis of gender inequality in the employment sector

Table of Contents

Gender Employment Parity Index (GEPI)

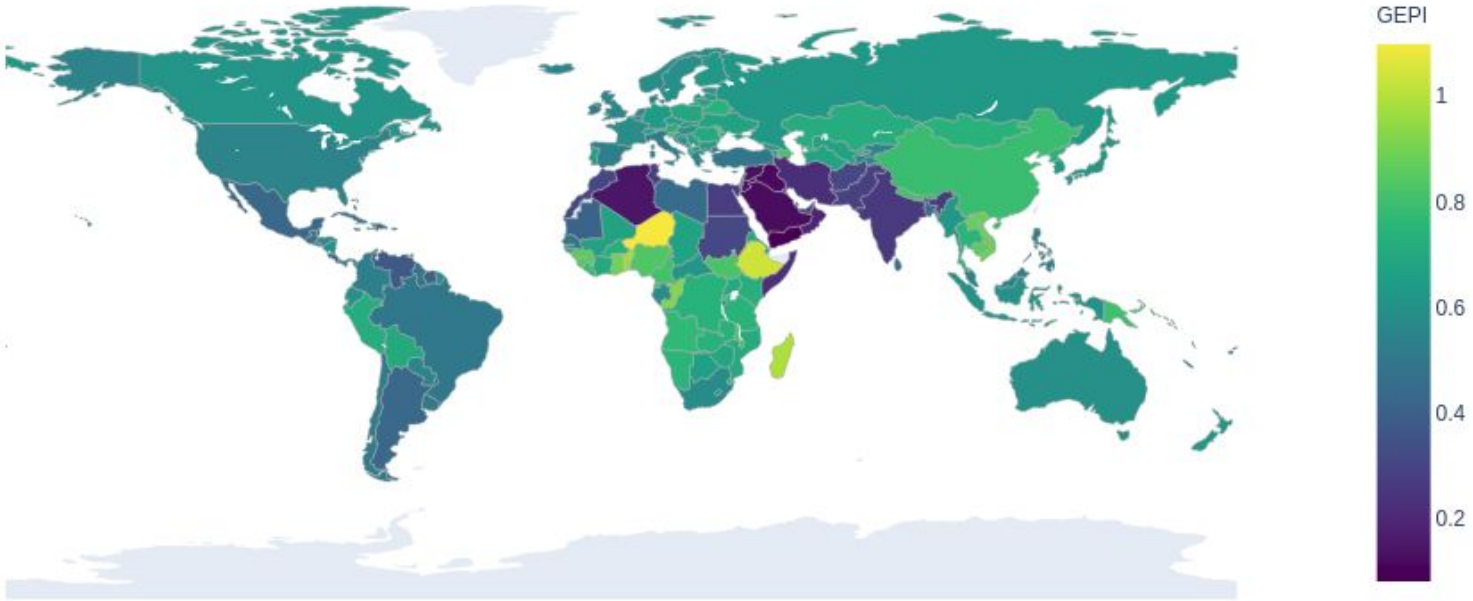
- Gender Parity Index (GPI) is a socioeconomic index introduced by UNESCO to identify gender bias in access to education.
- No index that can be used to define gender bias with respect to employment. Help formulate better policies
- This index can help in formulating government policies aimed at eradicating employment bias and providing equal job opportunities in each and every possible field irrespective of gender.

Gender Employment Parity Index (GEPI)

- For the data, a total of 187 countries spanning over 30 years (1991-2020) was considered.

- The GEPI was calculated as an average of three factors :
  - **Agriculture Employment Parity ratio**
  - **Industry Employment Parity ratio**
  - **Services Employment Parity ratio**

Average GEPI from 2016 to 2020



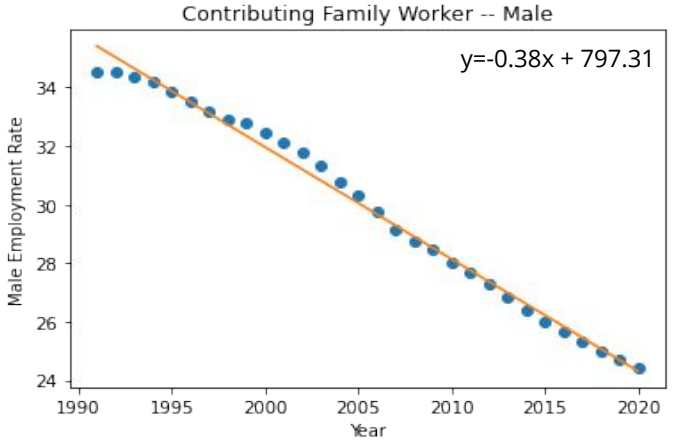
- 8 countries - 0-0.2 GEPI
- 17 countries - 0.2-0.4 GEPI
- 68 countries - 0.4-0.6 GEPI
- 73 countries - 0.6-0.8 GEPI
- 21 countries - >0.8 GEPI

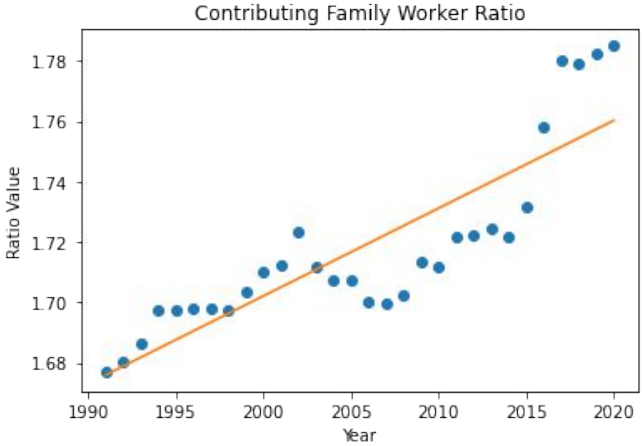
Table of Contents

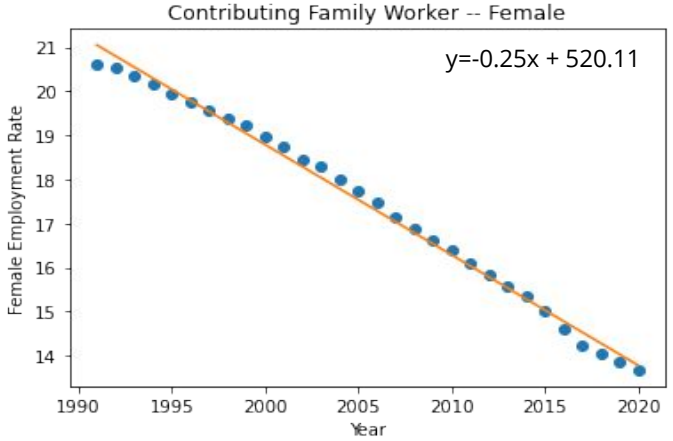
Contributing Family Workers

- Generally, male is more common to be the contributing family worker than the female.
- The employment rate is decreasing over the years.
- The contributing family worker ratio is increasing over the years.
- For 2021: 24.02% Male and 13.42% Female

$$\text{Ratio} = \frac{\text{Male Family Worker Rate}}{\text{Female Family Worker Rate}}$$

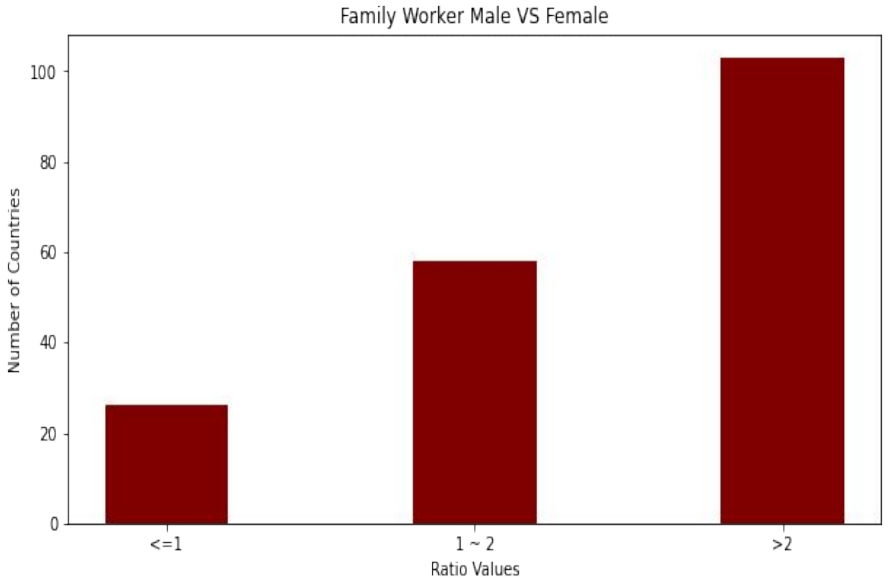






Contributing Family Worker

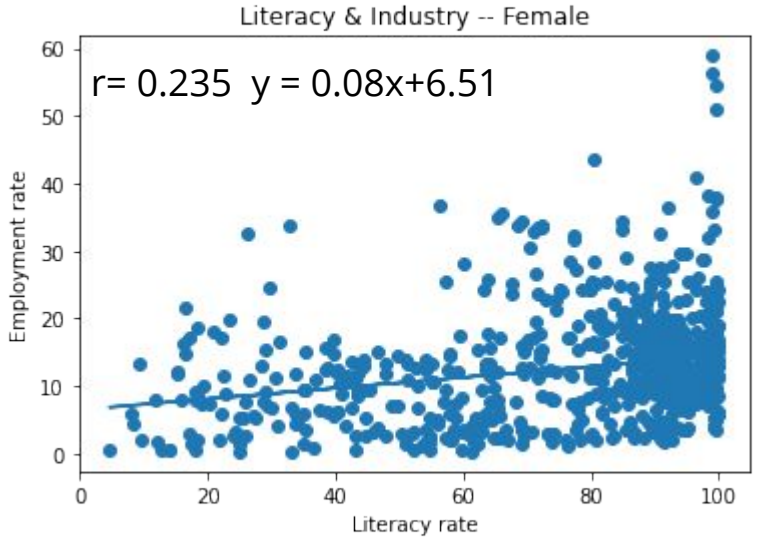
- We analyzed 187 countries and over than half of them have a ratio value greater than 2.
- Only 14% countries have a ratio less than 1 which means it is more common to see female as the contributing family worker.



$$\text{Ratio} = \frac{\text{Male Family Worker Rate}}{\text{Female Family Worker Rate}}$$

Table of Contents

Employment & Literacy



Employment_Rate_Industry_female

Country N	Country C	1991	1992
Afghanist	AFG	14.686	14.455
Angola	AGO	1.922	1.896
Albania	ALB	7.557	7.379
United An	ARE	14.606	14.432
Argentina	ARG	20.004	19.105
Armenia	ARM	10.155	9.833
Australia	AUS	12.444	12.259
Austria	AUT	18.095	17.863

Literacy_Rate_female

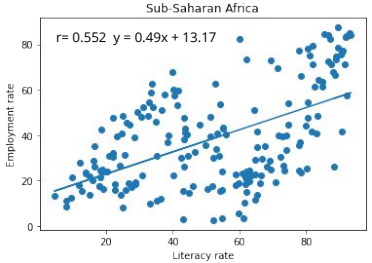
Country N	Country C	1991	1992
Aruba	ABW		
Afghanist	AFG		
Angola	AGO		
Albania	ALB		
Andorra	AND		
Arab Wor	ARB	41.9389	43.1267
United An	ARE		
Argentina	ARG	95.917	

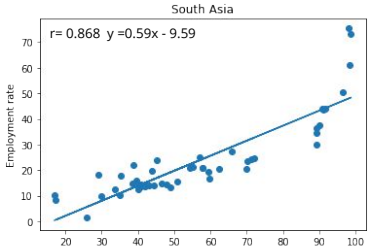
Datapoint Format:
Argentina_1991=(Employment_rate, Literacy_rate)
=(20.004, 95.917)

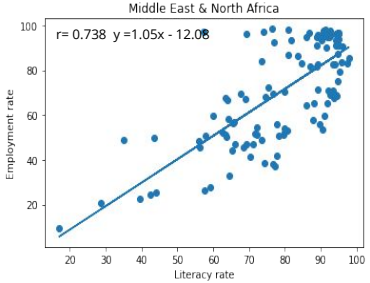
Services & Literacy

Positive Related

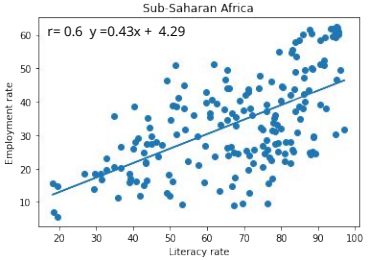
Female

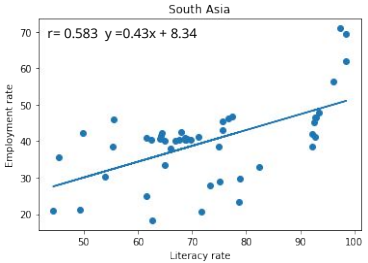


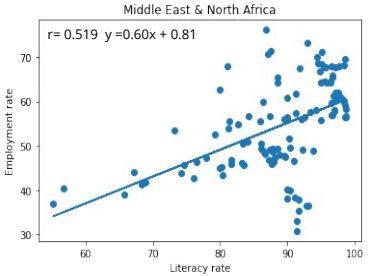


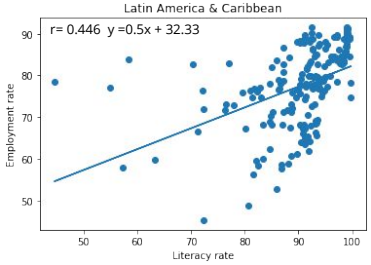


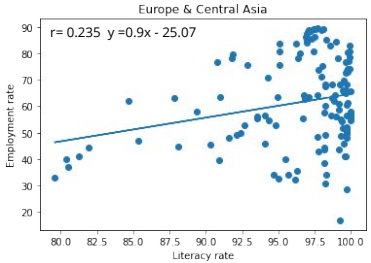
Male

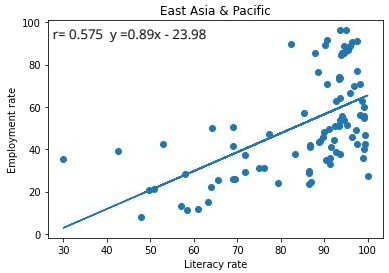


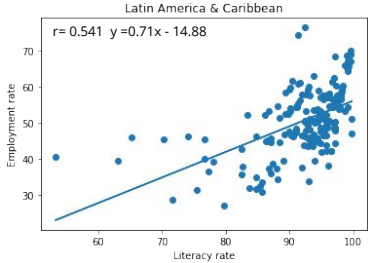


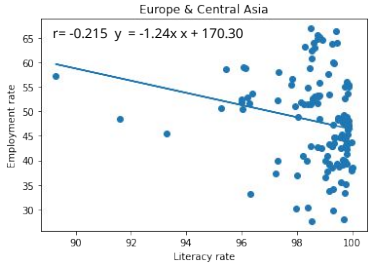


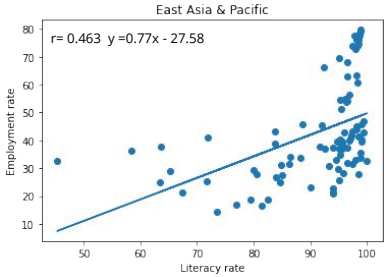




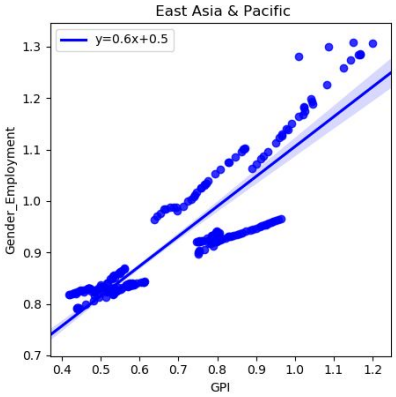


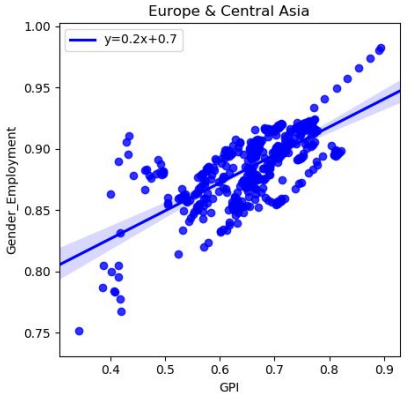


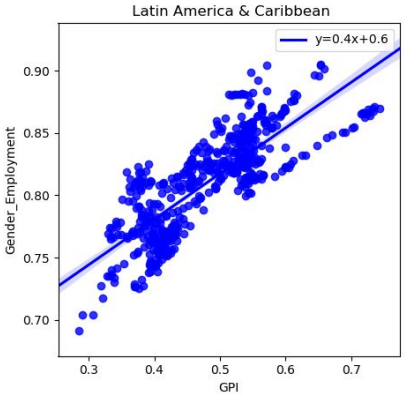


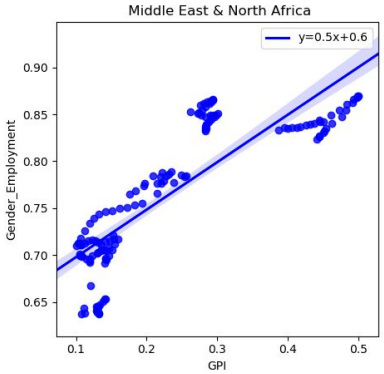


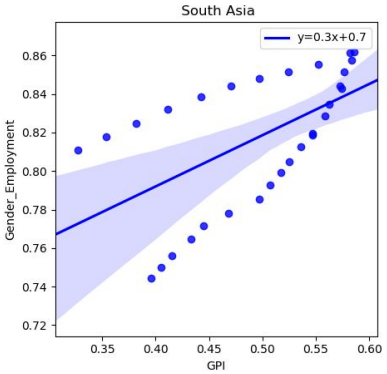
GPI vs Gender Employment

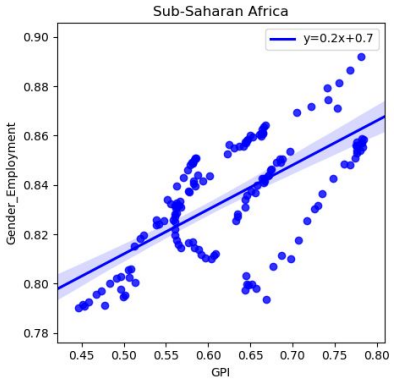




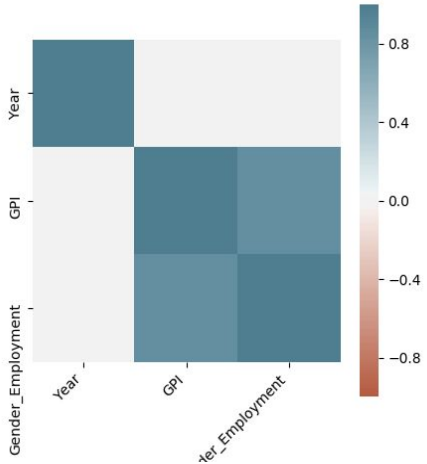


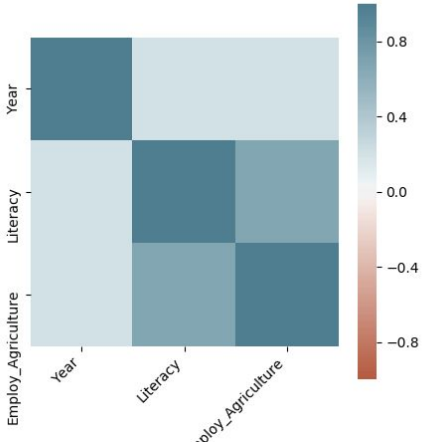


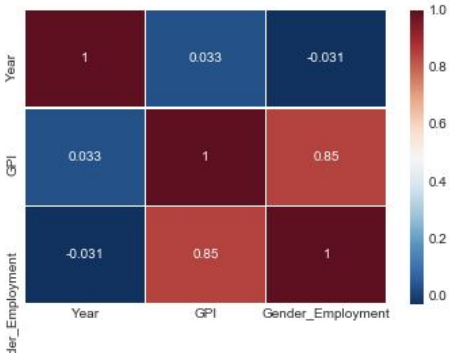




Correlation matrix:







The correlation matrix is used to verify the features which are strongly correlated. From the above plot it is evident that GPI and Gender employment are strongly correlated with a correlation coefficient of 0.85 and literacy vs employment is correlated with a correlation coefficient of 0.67. We have calculated the correlation grid using Pearson's coefficient.

Table of Contents

Time Series Prediction Models for Different Sectors of Employment

- We determined the employment ratios in three different sectors- Agriculture, Industry and Service for the next 5 years from 2021 to 2025.
- We then used these to calculate the corresponding GEPI.
- All these forecasts can help governments understand where their countries are headed towards in terms of reducing gender bias and formulate action plans accordingly.

Time Series Prediction Models for Different Sectors of Employment

- For each country, training and testing data are split in 83:17 ratio i.e. 25 years of data from 1991 to 2015 and 5 years of data from 2016 to 2020 respectively.
- Three metrics have been chosen for evaluating the model's performance in each employment sector:
  - Root mean square error (RMSE)
  - Mean average percentage error (MAPE)
  - Mean average error (MAE)
- We used four forecasting models for this purpose-
  - Simple Exponential Smoothing
  - Double Exponential Smoothing
  - ARIMA
  - Holt's Linear trend.

Time Series Prediction Models for Different Sectors of Employment

Table 1: Evaluation Metrics for the Four Models

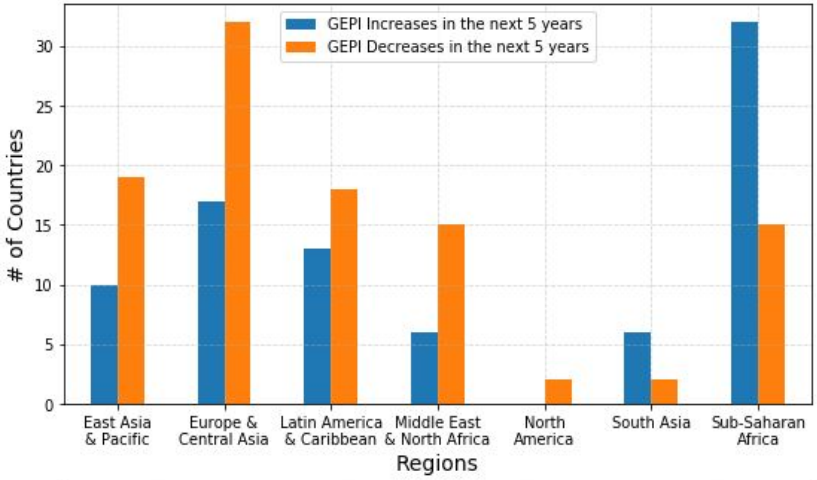
	Agriculture Sector			Industry Sector			Service Sector		
	RMSE	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*
SES Model	0.02716	0.07269	0.02553	0.01631	0.05122	0.01504	0.02786	0.03254	0.02592
DES Model	0.03353	0.08885	0.03063	0.02353	0.06839	0.02173	0.02904	0.03308	0.02622
ARIMA Model	0.03086	0.07921	0.02872	0.02086	0.06835	0.01903	0.02984	0.03755	0.02694
Holt's Model	0.03423	0.09440	0.03145	0.02235	0.06429	0.02058	0.02898	0.03313	0.02628

Table 2: Average of the Evaluation Metrics across all three factors

	RMSE	MAPE	MAE
SES Model	0.02378	0.05215	0.02216
DES Model	0.0287	0.06344	0.02619
ARIMA Model	0.02719	0.06170	0.02490
Holt's Model	0.02852	0.06394	0.02610

Analysis of Predicted GEPI based on the Different Regions and Income Groups

- European & Central Asian region - 65% GEPI decrease.
- Middle East & North African region - 71% GEPI decrease.
- South Asian region- 75% GEPI increase.
- Sub Saharan African region- 68% GEPI increase.
- High Income and Upper Middle income- GEPI decrease.
- Low Income and Lower Middle income- GEPI increase.



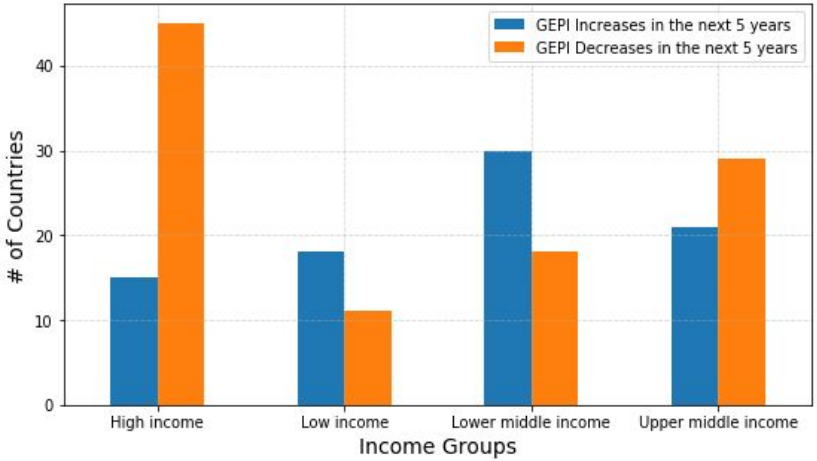


Table of Contents

Conclusion:

- Our project successfully defines a Gender Employment Index to recognize any gender bias.
- We also examined GEPI trends over the years with respect to income groups and regions the country belongs to.
- We successfully designed forecasting models to estimate GEPI in the future years.
- We performed correlation analysis between education and employment trends.

Table of Contents

Discussions & Future Work:

- How to handle the missing data
- Learned a lot about data visualization techniques and libraries used for it.
- Great exercise to write research papers and get introduced to the world of research.
- Find some more complete literacy dataset and expect to see a strong correlation between the employment rate and the literacy rate.
- Causality analysis between education and employment.
- Look into the gender bias with respect to manual and mental labor.
- Analyse the correlation between the role of women in a country's leadership positions like holding political offices, CEO of MNCs etc. and the GEPI ratios.

THANK YOU