Decoding : Inequality in Education

- Data (Kaggle) : https://www.kaggle.com/datasets/iamsouravbanerjee/inequality-in-education-around-the-world/data

- <mark>**Problem Statement:**</mark>

"AI-Driven Educational Equality Index: Identifying and Predicting Inequality Hotspots for Targeted Policy Interventions"

- Our goal is to:
1. **Measure and Rank Countries**: Develop an Educational Inequality Index (EII) that highlights the most affected regions.

2. **Predict Future Trends**: Build a predictive model to forecast inequality levels using historical trends.

3. **Find Key Determinants**: Identify the most critical factors driving inequality (HDI, economic status, region, etc.).

4. **Suggest Targeted Policies**: Provide AI-driven recommendations for policymakers to reduce disparities.

- How This Solution Benefits India & The World
  - **Targeted Education Reforms**: Helps governments focus resources on the most at-risk populations.
  - **Predictive Policymaking:** Forecasts inequality trends before they worsen.
  - **Global & Local Impact:** India, Africa, and low-HDI nations can use these insights to prioritize investments in education

- **Approach**
  - **Data Cleaning & Feature Engineering (Preprocessing)**
    - Handle missing values using interpolation.
    - Create an Educational Inequality Index (EII) as a composite metric.

  - **Exploratory Data Analysis (EDA)**
    - Visualize inequality trends by region & development level.
    - Analyze correlation with HDI, economic indicators.

  - **Predictive Model for Inequality**
    - Use Linear Regression to predict inequality levels for 2025 & beyond.
    - Identify the top 25 most at-risk nations.

  - **Policy Recommendations Using AI**
    - Implement clustering (K-Means) to group similar countries & tailor strategies.

Data Overview :

- **ISO3** - ISO3 for the Country/Territory

- **Country** - Name of the Country/Territory

- **Human Development Groups** - Human Development Groups

- **UNDP Developing Regions** - UNDP Developing Regions

- **HDI Rank (2021)** - Human Development Index Rank for 2021

- **Inequality in Education (2010)** - Inequality in Education for 2010

- **Inequality in Education (2011)** - Inequality in Education for 2011

- **Inequality in Education (2012)** - Inequality in Education for 2012

- **Inequality in Education (2013)** - Inequality in Education for 2013

- **Inequality in Education (2014)** - Inequality in Education for 2014

- **Inequality in Education (2015)** - Inequality in Education for 2015

- **Inequality in Education (2016)** - Inequality in Education for 2016

- **Inequality in Education (2017)** - Inequality in Education for 2017

- **Inequality in Education (2018)** - Inequality in Education for 2018

- **Inequality in Education (2019)** - Inequality in Education for 2019

- **Inequality in Education (2020)** - Inequality in Education for 2020

- **Inequality in Education (2021)** - Inequality in Education for 2021

Data Dictionary

- **UNDP Developing Regions**:

    - **SSA** - Sub-Saharan Africa

    - **LAC** - Latin America and the Caribbean

    - **EAP** - East Asia and the Pacific

    - **AS** - Arab States

    - **ECA** - Europe and Central Asia

    - **SA** - South Asia

ISO3	Country	Human Development Groups	UNDP Developeing Regions	HDI Rank (2021)	Inequality in Education (2010)	Inequality in edcuation (2011)	Inequality in Education (2012)	Inequality in Education (2013)	Inequality in Education (2014)	Inequality in Education (2015)	Inequality in Education (2016)	Inequality in Education (2017)	Inequality in Education (2018)
SOM	Somalia	NaN	AS	NaN	43.518790	43.518790	NaN	NaN	NaN	NaN	NaN	NaN	NaN
BIH	Bosnia and Herzegovina	High	ECA	74.0	21.478529	20.687452	20.687452	20.687452	19.888614	21.098021	16.974939	16.974939	14.788905
IRL	Ireland	Very High	NaN	8.0	3.231140	3.231140	3.231140	2.989530	2.931360	2.873080	2.873080	3.286040	3.286040
TCD	Chad	Low	SSA	190.0	41.861060	41.861060	41.861060	41.861060	41.861060	42.950100	42.950100	42.950100	42.950100
MCO	Monaco	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
FRA	France	Very High	NaN	28.0	9.435360	8.647990	7.961730	7.508710	8.648200	8.239880	8.596720	9.135000	9.486230
NPL	Nepal	Medium	SA	143.0	42.500000	41.377190	41.377190	41.377190	43.891120	43.891120	40.881000	40.881000	40.881000
MWI	Malawi	Low	SSA	169.0	30.201410	30.201410	30.201410	30.201410	28.213970	28.213970	28.424290	28.424290	28.424290
CAN	Canada	Very High	NaN	15.0	3.856000	3.856000	3.856000	2.678000	2.678000	2.696000	2.580000	2.521000	2.521000
KHM	Cambodia	Medium	EAP	146.0	28.329200	28.329200	28.329200	28.329200	27.304850	27.304850	27.304850	27.304850	27.304850

Dataset statistics

Number of variables	19
Number of observations	195
Missing cells	44
Missing cells (%)	1.2%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	29.1 KiB
Average record size in memory	152.7 B

Variable types

Text	2
Categorical	3
Numeric	14

Exploratory Data Analysis (EDA):

o **Statistics Summary :**

Overall EII Statistics:	
count	1.950000e+02
mean	-1.275333e-16
std	3.538119e+00
min	-4.850335e+00
25%	-2.890282e+00
50%	-5.421887e-01
75%	2.530024e+00
max	7.858402e+00

EII Statistics by UNDP Developing Regions								
UNDP Regions	count	mean	std	min	25%	50%	75%	max
AS	20.0	1.253464	2.739646	-2.831213	-0.593013	0.495328	2.755414	7.565894
EAP	26.0	-0.262274	2.280925	-3.143174	-1.779448	-0.433689	0.278567	7.094828
ECA	17.0	-2.901477	1.057106	-4.069685	-3.812826	-3.261592	-2.231467	-0.331923
LAC	33.0	-0.608325	2.052835	-3.591534	-2.039899	-0.932112	-0.075965	5.422730
SA	9.0	4.081763	3.802326	-3.573891	3.818031	5.461799	6.880584	7.143257
SSA	46.0	3.728896	2.499584	-0.828483	1.592338	3.768913	6.150443	7.858402

- **Key Takeaways:**
  1. **Sub-Saharan Africa (SSA):**

     Mean: 3.728896 (Highest among all regions)
     Median: 3.768913
     Standard Deviation: 2.499584

     Interpretation: SSA exhibits the highest average educational inequality. The median is close to the mean, suggesting a fairly symmetrical distribution of inequality levels within the region, but the substantial standard deviation indicates significant variation among countries in SSA. Some countries within SSA have much higher inequality than others.

  2. **South Asia (SA):**

     Mean: 4.081763 (Even higher than SSA)
     Median: 5.461799
     Standard Deviation: 3.802326 (Highest Standard Deviation)

     Interpretation: SA has even higher mean EII than SSA, indicating very severe educational inequality. Also note that the median (5.46) is substantially higher than the mean (4.08). This strongly suggests that the distribution of EII values within SA is skewed to the left. There are likely some countries with extremely high EII values pulling the median upwards. SA also has the highest standard deviation, confirming very high variability in inequality levels.

  3. **Arab States (AS):**

     Mean: 1.253464
     Median: 0.495328
     Standard Deviation: 2.739646

     Interpretation: The Arab States show moderate educational inequality compared to SSA and SA. The positive mean suggests inequality is a notable concern. The median is smaller than the mean, indicating that distribution is right-skewed, that means some countries have low educational inequality, and some have relatively high educational inequality.

  4. **East Asia & Pacific (EAP):**

     Mean: -0.262274
     Median: -0.433689

Standard Deviation: 2.280925

Interpretation: EAP shows relatively low educational inequality. The negative mean and median support the conclusion that inequality in this region is better than in other regions. The standard deviation suggests a degree of variability, but not as extreme as in SA or SSA.

5. **Latin America & Caribbean (LAC):**

   Mean: -0.608325
   Median: -0.932112
   Standard Deviation: 2.052835

   Interpretation: LAC presents low educational inequality, similarly to EAP. Also the same in EAP, the median is smaller than the mean, indicating that distribution is right-skewed, that means some countries have low educational inequality, and some have relatively high educational inequality

6. **Europe & Central Asia (ECA):**

   Mean: -2.901477 (Lowest Mean)
   Median: -3.261592 (Lowest Median)
   Standard Deviation: 1.057106 (Lowest Standard Deviation)

   Interpretation: ECA exhibits the lowest educational inequality. ==Both the mean and median are negative and have the lowest values, which indicates very little disparities in this region. The Standard Deviation is also the lowest, which means countries are very close to the mean value==, meaning there is relatively consistent equality across this region

- **Actionable Insights**
  1. **Prioritize SSA and SA:** These regions urgently require targeted interventions to address high levels of educational inequality. Specific strategies may include increasing access to schooling, improving education quality, and addressing gender disparities.

  2. **Understand Skewness:** The substantial difference between the mean and median in SA indicates skewed distributions. A deeper dive into country-specific data within SA is crucial to identify outlier countries with very high inequality and understand the underlying causes.

  3. **Leverage ECA's Success:** Europe & Central Asia serves as a positive model. ==Investigating the policies and practices in ECA that contribute to lower inequality could provide valuable insights for other regions==.

  4. **Contextualize Standard Deviation:** ==While a high standard deviation indicates variability, it's crucial to understand *why* variability exists. Are there specific economic, social, or political factors driving disparities within a region?==

5. **Address Skewed Data in Arab, LAC and EAP regions**: Conduct further studies on data skewness to better understand the reason.

| EII Statistics by Human Development Groups								
Human Development Groups	count	mean	std	min	25%	50%	75%	max
High	49.0	-1.083580	1.966241	-4.069685	-2.523567	-0.987337	-0.154708	4.617207
Low	32.0	5.147938	1.946411	0.709931	3.780087	5.536122	6.522037	7.858402
Medium	44.0	1.981058	2.667864	-3.261592	0.338796	1.743106	3.578116	7.094828
Unknown	4.0	-0.195252	0.476874	-0.433689	-0.433689	-0.433689	-0.195252	0.520060
Very High	66.0	-3.000366	1.499963	-4.850335	-4.397988	-3.383323	-2.001735	0.470595

- **Key Takeaways**
   1. **Very High Development:** Exhibits the lowest mean EII (-3.000366), indicating the least educational inequality. The relatively low standard deviation (1.499963) suggests consistency across countries in this group.

   2. **High Development**: Also shows a negative mean (-1.083580) and median (-0.987337), indicating low inequality.

   3. **Medium Development:** The mean is positive (1.981058), indicating moderate inequality. The standard deviation is the highest (2.667864), pointing to significant variability.

   Low Development: Shows the highest mean EII (5.147938), indicating severe educational inequality. This group needs the most urgent attention.

- **Graphical Summary :**

- **Illustration 1**

Inequality in Education Trends by Human Development Groups
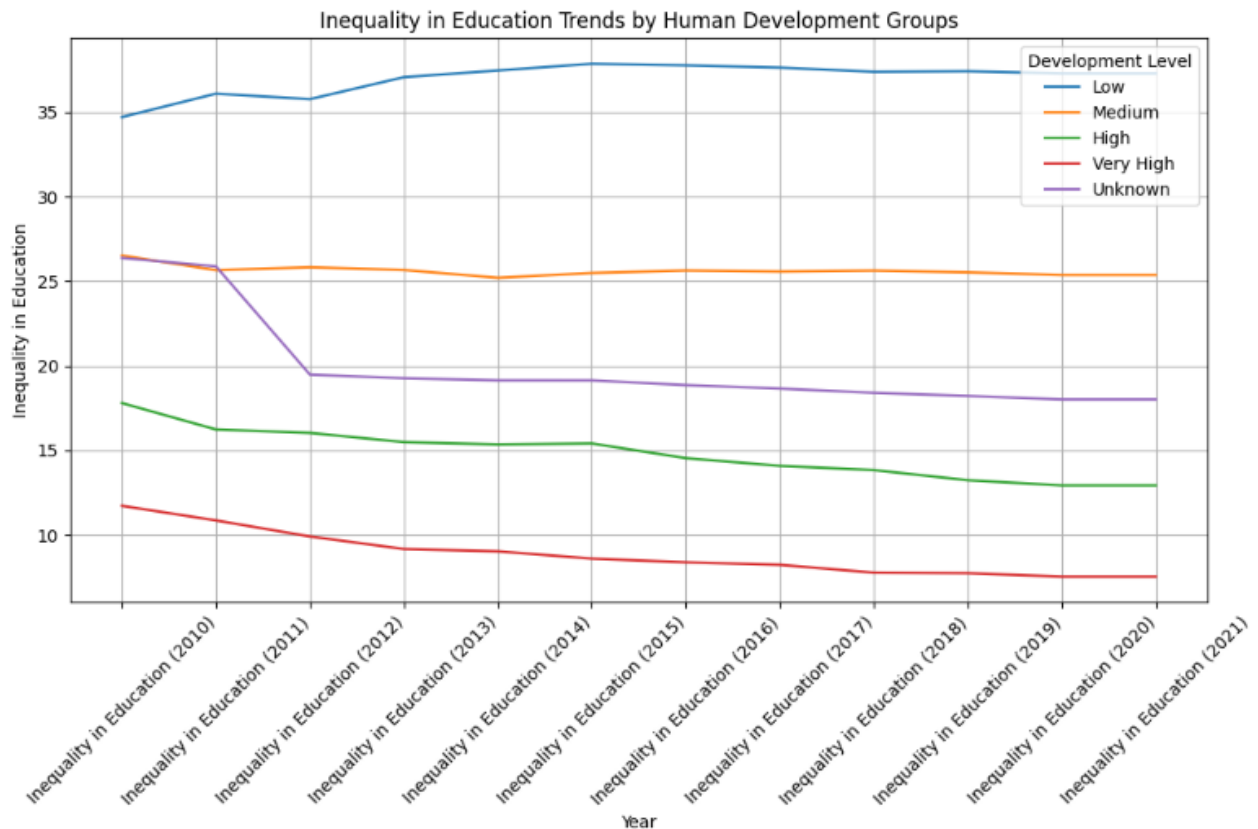
- **Explaination** :
  1. **Low Human Development Countries (Blue Line)**
     - Consistently have the **highest** education inequality.
     - Inequality increased slightly from 2010 to 2015 and then remained relatively stable.

  2. **Medium Human Development Countries (Orange Line)**
     - Show a **steady** trend with minor fluctuations.
     - No significant improvement in reducing inequality.

  3. **High Human Development Countries (Green Line)**
     - **Gradually decreasing** inequality over time.
     - Suggests efforts in education access and quality improvements.

  4. **Very High Human Development Countries (Red Line)**
   - **Lowest education inequality** among all groups.
     - Shows a **consistent decline**, indicating strong educational policies reducing disparity.

  5. **Unknown Development Level (Purple Line)**
     - Started at a mid-level inequality but saw a **sharp drop around 2011**, stabilizing afterward.
   - This could indicate data reclassification or major policy shifts.

- **Key Takeaways:**
  - Higher development levels correlate with lower education inequality.
  - Inequality is persistent in low and medium development regions.
  - Very high development countries show continuous improvement**.**

    o **Illustration 2**

Correlation between HDI Rank and Inequality in Education
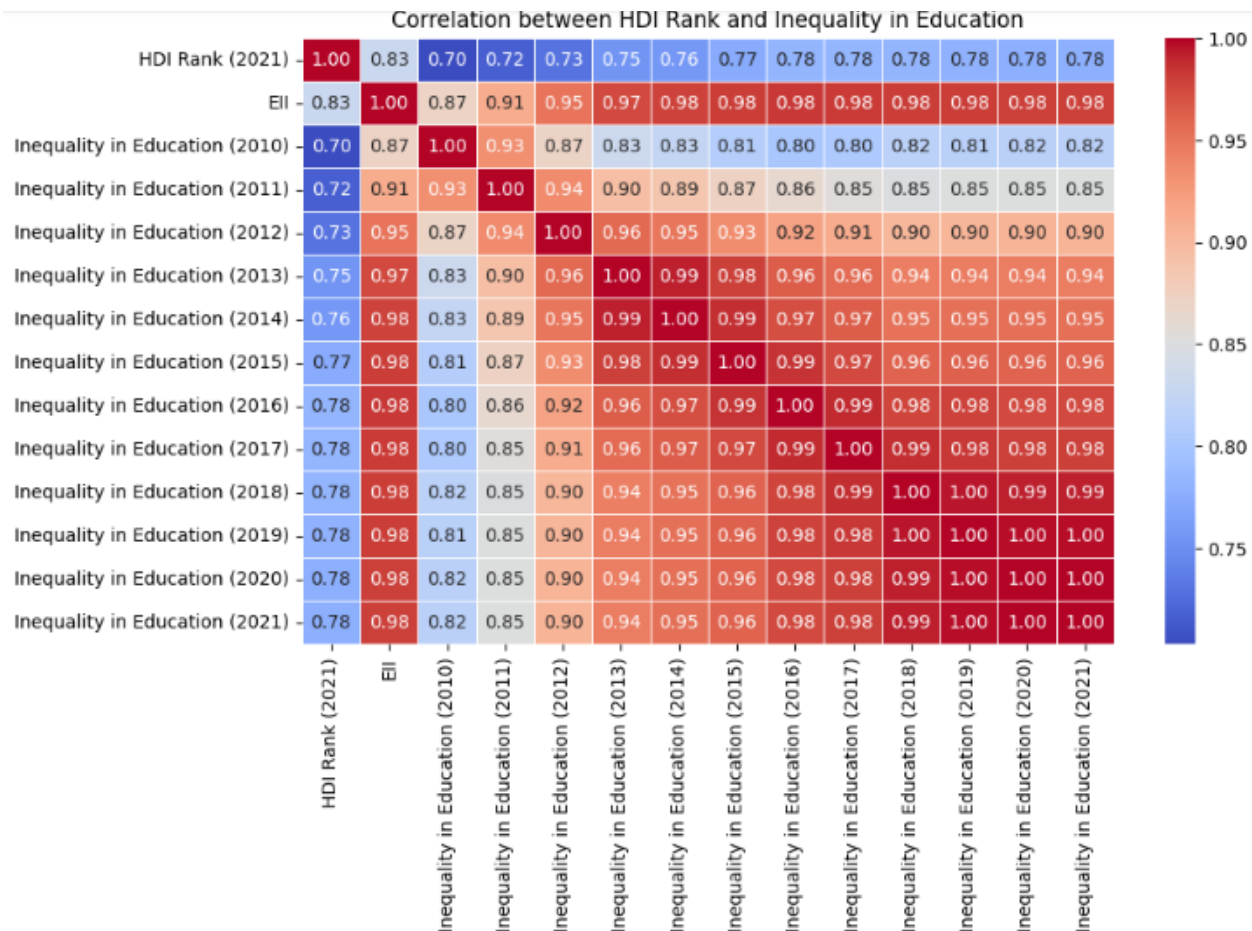
- **Key Takeaways :**
  1. HDI rank is negatively correlated with inequality in education across all years (2010–2021). **As HDI ranks improves (lower numerical values)** are associated with lower inequality in education.
  2. **Trend**: The correlations are consistently strong, ranging from approximately -0.72 to -0.78, indicating a stable relationship over time.

2x2 Matrix: HDI Rank vs. Average Education Inequality
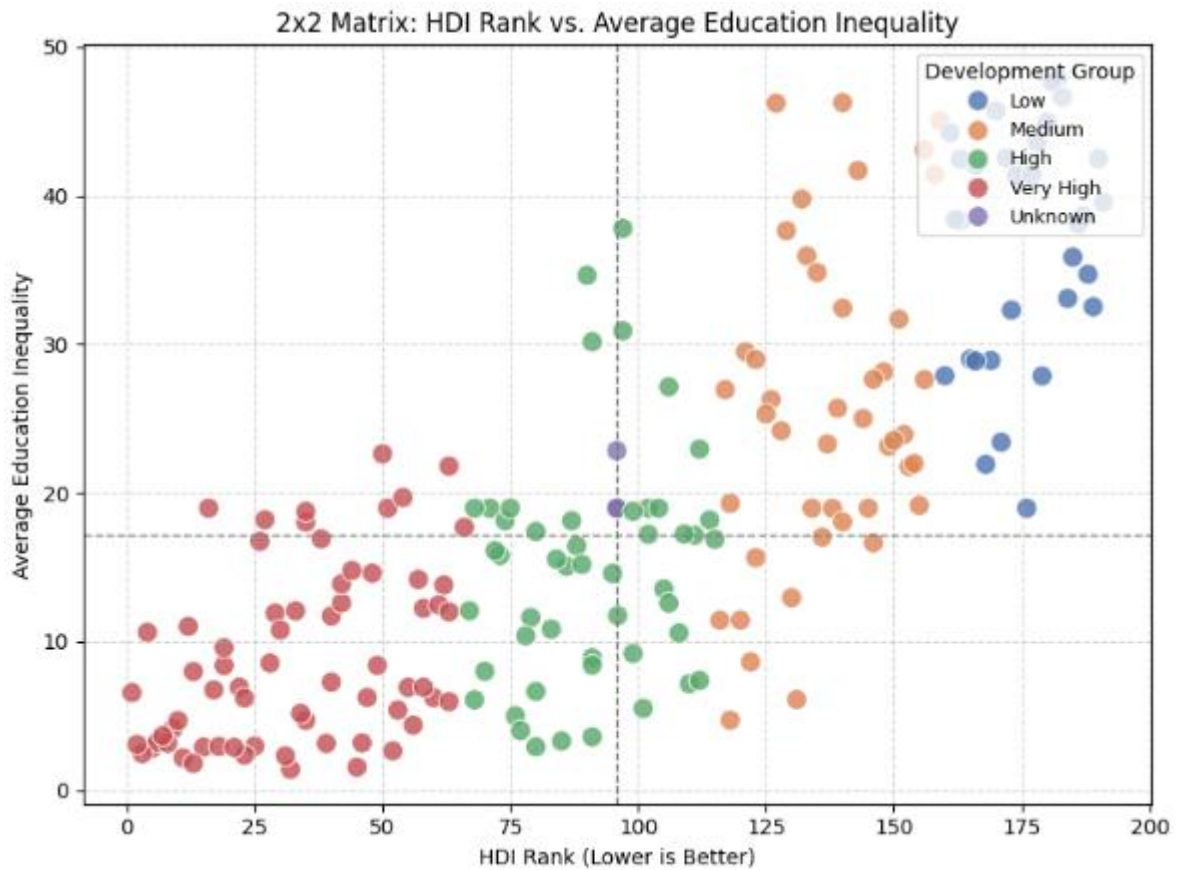
- **Key Takeaways :**
    1. **Cluster Analysis:**
        a. **Very High Development Group**: Concentrated near low HDI ranks and low education inequality.
        b. **Low Development Group**: Found near high HDI ranks and high education inequality.
        c. **Medium and High Development Groups:** Spread across intermediate ranges.
    2. **Conclusion:** Development level strongly influences both HDI rank and education inequality.

❖ **Illustration 4**

Sankey Diagram: Human Development Groups to UNDP Developing Regions

- **Key Takeaways :**
    1. The distribution of development groups varies significantly across UNDP regions:
        a. **Sub-Saharan Africa (SSA):** Predominantly consists of "Low" and "Medium" development countries, reflecting higher inequality.
        b. **Europe & Central Asia (ECA):** Dominated by "High" and "Very High" development groups, with lower inequality.
        c. **South Asia (SA):** Predominantly "Low" development countries, similar to SSA.
        d. **East Asia & Pacific (EAP) and Latin America & the Caribbean (LAC):** Mixed distribution across all development groups.

    Conclusion: Inequality in education correlates strongly with human development levels, with SSA and SA regions requiring significant interventions

Average Educational Inequality Index by UNDP Region
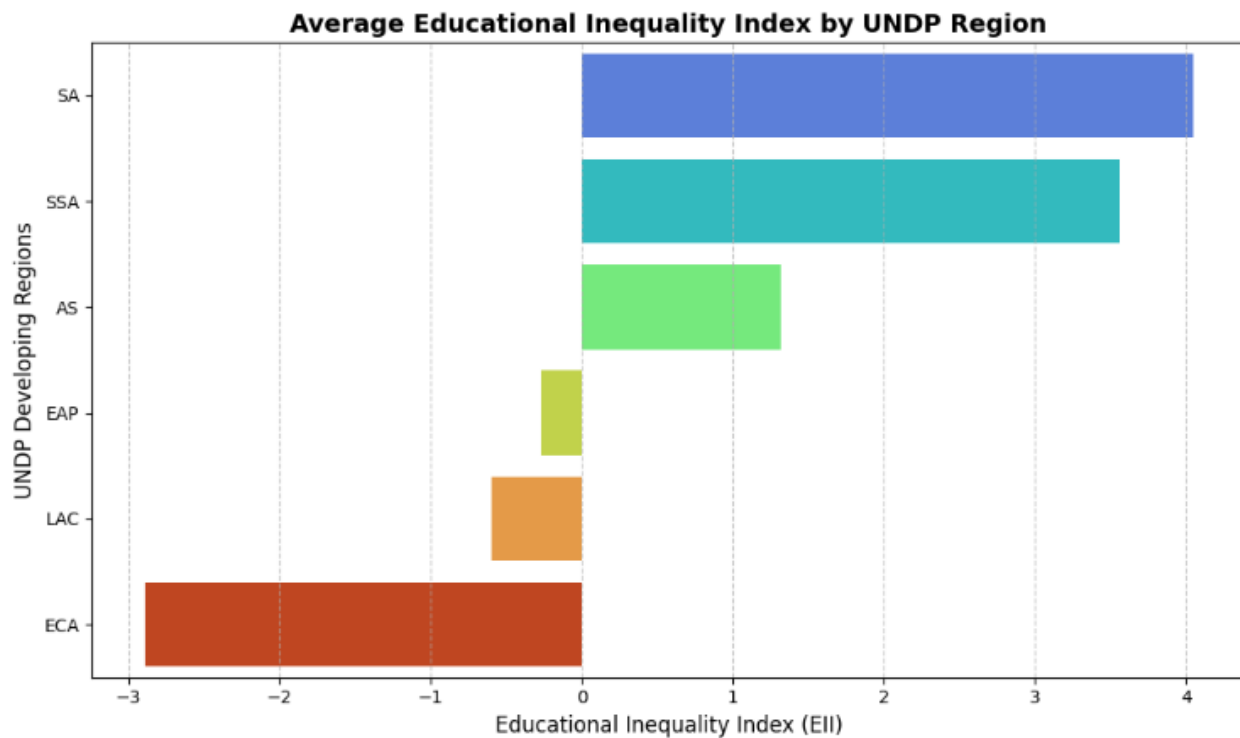
- **Key Takeaways :**
    1. **SSA and SA:** Have the highest <mark>average educational inequality index (EII)</mark> , highlighting significant disparities.

    2. **ECA**: Displays the lowest inequality index, indicating better access to education across populations.

    3. Other regions like LAC and EAP show intermediate levels of inequality.

    Conclusion: SSA and SA are the most vulnerable regions regarding educational inequality, while ECA leads in equitable education outcomes.

❖   **Overall Analysis Findings:**

	Key Findings	Interpretation & Implications	Policy Recommendations / Insights
Regional Analysis	• **Sub-Saharan Africa (SSA)**: Mean = 3.73, Median = 3.77, SD = 2.50 • **South Asia (SA):** Mean = 4.08, Median = 5.46, SD = 3.80 • **Arab States (AS):** Mean = 1.25, Median = 0.50, SD = 2.74 • **EAP & LAC:** Negative means, moderate inequality • **ECA**: Lowest inequality (Mean = -2.90, Median = -3.26, SD = 1.06)	• SSA and SA exhibit the highest average educational inequality, with SA showing a highly skewed and variable distribution. • ECA, by contrast, has very low and consistent inequality levels. • Moderate levels in AS, EAP, and LAC indicate mixed outcomes	• Prioritize targeted interventions in SSA and SA. • Investigate and adapt successful policies from ECA to other regions. • Conduct deeper analysis for regions with skewed distributions (e.g., SA) to understand outlier country dynamics

Development Group Analysis	• **Low Development:** Highest mean EII (severe inequality) • **Medium Development:** Positive mean with highest variability • **High & Very High Development:** Negative means, indicating lower inequality with consistency	• A clear gradient is evident: lower development correlates with higher inequality, while very high development groups show the least inequality. • The medium development group exhibits significant variability, suggesting uneven progress within this category.	• Direct focus on low development countries for immediate interventions. • Tailor strategies for medium development countries to address variability. • Leverage best practices from high/very high development groups
Time Trend Analysis	• **Low HD Countries:** Highest and persistent inequality. • **Medium HD Countries:** Steady with minor fluctuations. • **High HD Countries**: Gradual decrease in inequality. • **Very High HD Countries**: Consistent decline in inequality over time	• Trends over time reveal that while high and very high development groups are improving, low and medium groups remain stagnant or only slightly improve. • The persistent high inequality in low and medium groups underlines systemic issues that have not been effectively addressed over time.	• Accelerate interventions in low and medium development regions with focused educational and socio-economic programs. • Identify successful policies in high/very high groups and adapt them for low-performing regions.
Correlation Analysis (HDI vs. EII)	• Strong negative correlation (ranging approximately from -0.72 to -0.78) between HDI rank and educational inequality across all years.	• A negative correlation indicates that better HDI performance (lower rank values) is consistently associated with lower educational inequality. • The stable and strong relationship across years underscores the reliability of HDI as a proxy for overall development influencing education equality.	• Integrate policies that simultaneously improve HDI components (e.g., healthcare, income, education) to indirectly lower educational inequality. • Use HDI as a benchmark for tracking policy impact over time
Cluster& Distribution Analysis	•**Cluster Analysis:** - Very High Development clusters near low HDI ranks and low in equality. - Low Development clusters near high HDI ranks and high inequality. • Distribution: Varies significantly by region (e.g., SSA and SA dominated by low/medium groups)	• The clusters clearly show that development level is a strong determinant of educational inequality. • Variability and skewness in distributions (especially in SA and some Arab States) indicate the presence of outlier countries with extreme inequality that require special attention.	• Design region- and cluster-specific interventions. • For regions with significant outliers, conduct country-level analyses to tailor policy responses to unique local challenges.

Predictive Model for Inequality

1. **Model : Linear Regression** (Y=β0+β1×X)
   (Not used ARIMA or LSTM model as there is only 12 data points for each to predict the value of y)

Ex – For Afghanistan ,
 X = [2010,2011,2012,…,2021] (years)
 Y = [42.809,44.82338,44.82338,…,45.36517]Y = [42.809, 44.82338, 44.82338, …, 45.36517]Y=[42.809,44.82338,44.82338,…,45.36517] (corresponding inequality values for Afghanistan)

Then calculated **constant** and **beta coefficient** value **: β1 ≈0.14378 , constant : −244.81**

$$y(2025) = -244.81 + 0.14378 \times 2025 = 46.34 \text{ (predicted value for 2025)}$$

> Predicted Inequality Levels for 2025, 2030, 2035 (here, stated 25 countries including India)

Index	Country	2025	2030	2035
0	Afghanistan	46.337461	47.056356	47.775252
75	Haiti	35.59406	34.17279	32.751513
76	Hungary	2.559292	2.227494	1.8956965
77	Indonesia	14.86437	13.11714	11.369903
78	India	33.81367	30.67183	27.529986
79	Ireland	3.374975	3.484088	3.5932017
80	Iran	4.932319	4.896365	4.8604104
81	Iraq	32.20908	33.63249	35.055902
82	Iceland	2.173133	2.023219	1.8733042
83	Israel	5.007881	3.997364	2.9868475
84	Italy	9.127741	8.258716	7.3896914
85	Jamaica	2.631893	0.246958	-2.1379772
86	Jordan	11.85991	9.044167	6.2284258
87	Japan	-5.6124	-12.9838	-20.355187
88	Kazakhstan	0.846532	-1.01506	-2.8766461
89	Kenya	20.61744	18.87067	17.123903
90	Kyrgyzstan	1.915344	0.441729	-1.0318856
91	Cambodia	26.55748	25.98442	25.411355
92	Kiribati	4.188815	-2.56835	-9.3255128
93	Saint Kitts and Nevis	16.92584	15.84415	14.762463
94	South Korea	7.492618	6.39059	5.2885618
95	Kuwait	20.95457	20.0673	19.180019
96	Lao	31.14844	30.46862	29.788797
97	Lebanon	2.093625	-0.69049	-3.4746126
98	Liberia	39.97677	38.05771	36.138657
99	Libya	16.92584	15.84415	14.762463

> **Top 25 Most At-Risk Nations in 2025**

| index	Country	2025	2030	2035	Average Forecast
65	Gambia	65.3144317	77.9341962	90.5539607	77.9341962
115	Mali	60.2212933	71.85851829	83.49574328	71.85851829
38	Comoros	61.17438424	70.70519867	80.2360131	70.70519867
14	Burkina Faso	52.42164872	62.59585033	72.77005194	62.59585033
64	Guinea	53.97875303	57.20929848	60.43984394	57.20929848
140	Papua New Guinea	46.47055569	56.38664553	66.30273538	56.38664553
13	Benin	49.69444911	53.72362763	57.75280615	53.72362763
1	Angola	44.71338507	53.43473211	62.15607915	53.43473211
120	Mozambique	47.04117651	52.92208469	58.80299288	52.92208469
55	Ethiopia	48.7644424	52.44935117	56.13425994	52.44935117
26	Bhutan	50.20525236	52.2954042	54.38555604	52.2954042
107	Morocco	42.99238021	50.36644557	57.74051094	50.36644557
152	Senegal	48.49424097	49.97312086	51.45200076	49.97312086
49	Algeria	41.97369985	48.18331732	54.3929348	48.18331732
33	Ivory Coast	46.90654056	47.88805664	48.86957273	47.88805664
126	Niger	42.22954537	47.33744022	52.44533507	47.33744022
0	Afghanistan	46.33746086	47.05635649	47.77525212	47.05635649
155	Sierra Leone	46.45077934	45.82275399	45.19472864	45.82275399
169	Chad	43.74271166	44.4009634	45.05921515	44.4009634
66	Guinea-Bissau	43.27676848	44.25264576	45.22852303	44.25264576
121	Mauritania	43.08623484	43.96055931	44.83488379	43.96055931
191	Yemen	44.87468864	43.95306818	43.03144773	43.95306818
151	Sudan	42.28883625	42.16509918	42.04136212	42.16509918
135	Pakistan	42.51694055	41.61428338	40.71162621	41.61428338
174	Timor-Leste	43.02481077	41.31799538	39.61118	41.31799538

Clustering the countries based on their average inequality value

- Implement clustering (K-Means) to group similar countries & tailor strategies.

index	0	1	2	3
0	Angola	Andorra	Afghanistan	Albania
1	Bolivia	United Arab Emirates	Burundi	Armenia
2	Botswana	Argentina	Benin	Antigua and Barbuda
3	Cameroon	Australia	Burkina Faso	Azerbaijan
4	The Democratic Republic of the Congo	Austria	Bangladesh	Bulgaria
5	Congo	Belgium	Bhutan	Bosnia and Herzegovina
6	Cabo Verde	Bahrain	Central African Republic	Belize
7	Djibouti	Bahamas	Ivory Coast	Brazil
8	Algeria	Belarus	Comoros	Barbados
9	Egypt	Canada	Ethiopia	Brunei
10	Eritrea	Switzerland	Ghana	China
11	Micronesia	Chile	Guinea	Colombia
12	Gabon	Cyprus	Gambia	Costa Rica
13	Equatorial Guinea	Czechia	Guinea-Bissau	Cuba
14	Honduras	Germany	Guatemala	Dominica
15	Iraq	Denmark	Haiti	Dominican Republic
16	Kenya	Spain	India	Ecuador
17	Cambodia	Estonia	Liberia	Fiji
18	Kiribati	Finland	Madagascar	Grenada
19	Lao	France	Mali	Guyana
20	Saint Lucia	United Kingdom	Mozambique	Indonesia
21	Lesotho	Georgia	Mauritania	Iran
22	Morocco	Greece	Niger	Jamaica

23	Maldives	Hong Kong	Nigeria	Jordan
24	Myanmar	Croatia	Nepal	Kyrgyzstan
25	Malawi	Hungary	Pakistan	Saint Kitts and Nevis
26	Namibia	Ireland	Sudan	Kuwait
27	Nicaragua	Iceland	Senegal	Lebanon
28	Papua New Guinea	Israel	Sierra Leone	Libya
29	Rwanda	Italy	South Sudan	Sri Lanka
30	Solomon Islands	Japan	Chad	Monaco
31	El Salvador	Kazakhstan	Togo	Moldova
32	Sao Tome and Principe	South Korea	Timor-Leste	Mexico
33	Eswatini	Liechtenstein	Yemen	Marshall Islands
34	Syrian Arab Republic	Lithuania		North Macedonia
35	Tunisia	Luxembourg		Mongolia
36	Tanzania	Latvia		Mauritius
37	Uganda	Malta		Malaysia
38	Vanuatu	Montenegro		Nauru
39	Zambia	Netherlands		Oman
40	Zimbabwe	Norway		Panama
41		New Zealand		Peru
42		Poland		Philippines
43		Portugal		Palau
44		Qatar		North Korea
45		Romania		Paraguay
46		Russian Federation		Palestine, State of
47		Saudi Arabia		Somalia
48		Singapore		Serbia
49		San Marino		Suriname
50		Slovakia		Seychelles
51		Slovenia		Thailand

52	Sweden	Tajikistan
53	Turkey	Turkmenistan
54	Uruguay	Tonga
55	United States	Trinidad and Tobago
56		Tuvalu
57		Ukraine
58		Uzbekistan
59		Saint Vincent and the Grenadines
60		Venezuela
61		Viet Nam
62		Samoa
63		South Africa

K-Means Clustering of Countries
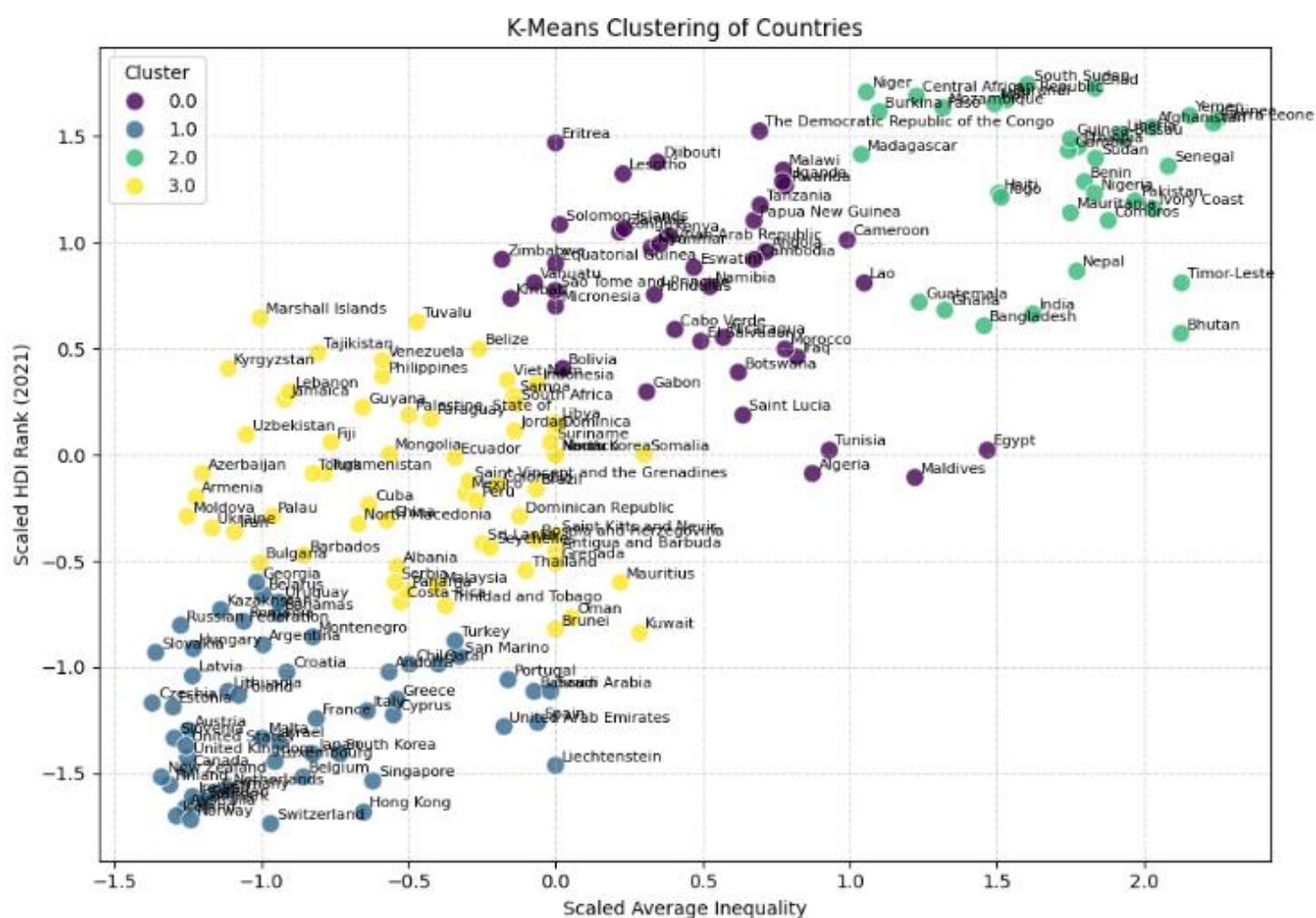
- Countries in Cluster 2 exhibit the highest average levels of inequality in education compared to other clusters.

SHAP Feature Importance for Predicting Educational Inequality
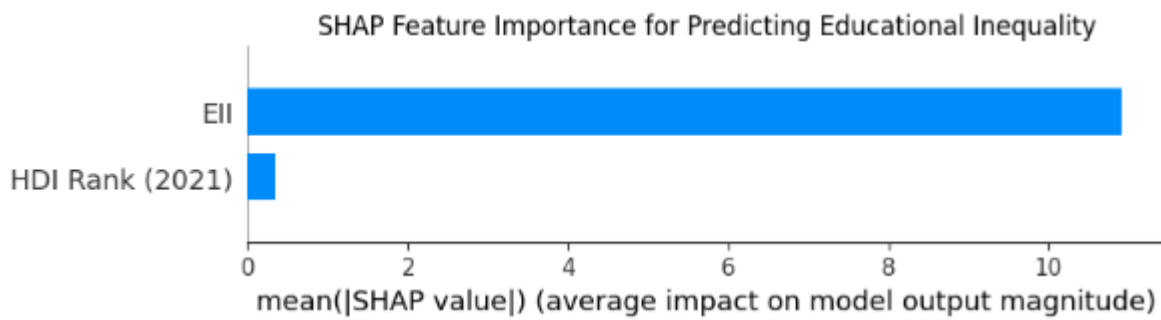
Recommended Policies for United Nations to Address Educational Inequality:

1. Prioritize Interventions in High-Inequality Regions

Sub-Saharan Africa (SSA) and South Asia (SA) exhibit the highest levels of educational inequality. These regions require urgent and focused interventions:

- **SSA:** Develop programs to expand access to education, improve infrastructure, and address gender disparities. Focus on rural and marginalized communities where inequality is most pronounced.

- **SA:** Conduct country-specific analyses to understand the skewed distribution of inequality. Policies should target outlier countries with extreme disparities in education access and quality.

2. Leverage Successful Practices from Europe & Central Asia (ECA)

ECA demonstrates the lowest levels of educational inequality due to consistent policies and equitable distribution of resources. The UN can:

- Study and adapt ECA's policies related to universal access, quality improvements, and financial support for education.

- Promote cross-regional collaboration to share best practices with SSA, SA, and other regions facing challenges.

3. Address Variability in Medium Development Countries

The medium development group shows significant variability in educational inequality. Tailored strategies are necessary:

- Identify specific socio-economic or political factors causing disparities within this group.

- Implement region-specific programs that address variability while promoting consistency in education outcomes.

4. Accelerate Interventions for Low Development Countries

Countries with low human development exhibit the highest educational inequality. Immediate actions should include:

- Increasing investments in primary and secondary education infrastructure.

- Providing subsidies or free education programs for underprivileged communities.

- Strengthening teacher training programs to improve education quality.

5. Predictive Policymaking Using AI

Utilize AI-driven models to forecast future trends in educational inequality and identify hotspots for intervention:

- Build predictive algorithms based on historical data to anticipate worsening inequality levels.

- Use clustering techniques (e.g., K-Means) to group countries with similar challenges and tailor strategies accordingly.

6. Promote Gender Equality in Education

Gender disparities are a major driver of inequality, particularly in SSA and SA. Policies should focus on:

- Increasing enrollment rates for girls through scholarships, awareness campaigns, and community engagement.

- Addressing cultural barriers that prevent girls from accessing education.

  .

7. Contextualize Data Skewness

Regions like SA, Arab States (AS), Latin America & Caribbean (LAC), and East Asia & Pacific (EAP) show skewed distributions of inequality. Deeper analysis is needed:

- Investigate outlier countries within these regions to understand specific causes of skewness.

- Develop tailored interventions targeting countries with high inequality while maintaining progress for better-performing nations.

8. Strengthen Correlation Between HDI and Education

The strong negative correlation between HDI rank and educational inequality highlights the importance of socio-economic factors:

- Integrate education reforms with broader socio-economic development programs like poverty alleviation, healthcare access, and gender equality initiatives.

- Promote multi-sectoral collaboration among governments, NGOs, and international organizations

  .

9. Focused Time-Trend Interventions

While high and very high development groups show improvement over time, low and medium groups remain stagnant:

- Implement long-term policies targeting systemic issues such as funding gaps, teacher shortages, and infrastructure deficits.

- Monitor progress using time-trend analyses to ensure sustained improvements.

10. Global Advocacy for Educational Equity

The UN can lead global advocacy campaigns emphasizing the importance of equitable education as a fundamental human right:

- Encourage member states to commit resources toward reducing disparities.

- Facilitate international partnerships to share resources, expertise, and funding for educational reforms.

Conclusion

The proposed policies aim to reduce educational inequalities by focusing on regions with severe disparities while leveraging successful practices from better-performing areas. By integrating AI-driven insights with targeted interventions, the United Nations can play a pivotal role in ensuring equitable access to education worldwide.