World Bank Education Analysis

Exploratory Data Analysis

Prepared by : Azad Meshram

Data science trainee,

AlmaBetter, Bangalore

Abstract:

Humans are at the pedestal of their development journey. This stage of culmination is achieved by the extensive and long process of learning things about our surroundings and getting educated. The foremost important challenge in front of education is its applicability and its transformation. More the world facilitates in future, new education is required. The historical studies performed by the World Bank would help humanity to fetch the itinerary for the requirement of new world.

Introduction:

The World Bank Ed Stats All Indicator Query holds over 4,000 internationally comparable indicators that describe education access, progression, completion, literacy, teachers, population, expenditures and many such indicators .The indicators cover the education cycle from pre-primary to vocational and tertiary education and also holds learning outcome data from international and regional learning assessments (e.g. PISA, TIMSS, PIRLS), equity data from household surveys, and projection/attainment data.

I am exploring and analyzing the data to identify variation of indicators across the globe, which countries are more alike and different, which countries show more closeness to global standards and what are the actual driving force behind such growth.

Data Exploration

There are five data sets given by World Bank : But there was only one dataset on which World Bank have worked. It contains 886390 rows The details of attributes of Ed Stats data set is given below:

Columns	Description
Country Name	Name of the Countries
Country Code	Abbreviation for the Country name
Indicator Name	The head on which trends are observed.
Indicator Code	The code form of a Indicator.
Years	Form 1970-2100

In indicator file , World Bank have studied 3666 indicators over 241

countries giving there trends from 1970 to present years.
The Data frame is full of null values and devoid of duplicate values.
In the Ed Stats Dataset there are 886390 rows and 70 columns. The Country Name , Country Code , Indicator Name and Indicator code has 'object' as their type. The Year Columns (1970-2100) has the float values.

Standard Operating Procedure:

In the process of Exploratory Data Analysis I followed a particular procedure for obtaining best outcome from the data:

1. Importing required packages for analysis.
2. Mounting drive and reading data files from Google drive.
3. Viewing all data information.
4. Checking unique values, null count and datatypes of each column.
5. Identifying the problem statements.
6. Doing concatenation, groupby , sorting and aggregating operation to come up to relevant data frame.
7. Doing data visualization as per the problem statements.
8. Finding Insights from the data visualization.
9. Giving recommendations.

.Data Cleaning Operation:

1. To observe the probable trends, I tried to make full out of the given dataset.
2. I replaced the null values with zero.
3. With this the Year would not be dropped from the comparative analysis data frame of more than one indicator at once.
4. Later on When I need to find out the mean of newly formed desired data frames , I replaced those zeros back with their respective null values . So that the mean should not be affected.

Data frames Used:

In the process of data analysis I formed many data frames to breakdown the given data set , list of those are given below:

Data Frames	Description
first_df	Ed Stats data
Second_df	Ed Stats Country
Arab_df	Trends of four indicators of Arab World
china_df	Trends of four indicators of Arab World
UK_df	Trends of three indicators of United Kingdom
Arg_df	Trends of indicators of Argentina
Japan_df	Trends of indicators of Japan
ZAM_df	Trends of indicators of Zambia
BOTS_df	Trends of indicators of Botswana
EAP_df	Trends of indicators of East Asia Pacific
GER_df	Comparing Gross Enrollment of countries.
LSR_df	Comparing Lower Secondary completion rate
UN_df	Comparing Unemployment
GEE_four	For the pupil teacher trends of all countries.
GEE_four	Comparing Govt Expenditure on Education
PTR_df	For the pupil teacher trends of all countries.
world_df	For the all the trends of all countries

User Defined Functions used:

1. **Data_Extracter(Country Name, Indicator Name):**
   The function used for the Data Extraction of all the countries, using the name of country and name of indicator.
2. **Concat(*variable):**
   This function takes the output data frames as its argument, and concat them at axis 1.
3. **Unstacker(*indicator):**
   This function unstack the desired data frame for easy plotting.
4. **Present_values(df , country):**
   This function returns a beautiful data frame which is unstacked first and then groupby to calculate the mean of country specific values.

Approach used:

Exploratory Data Analysis:

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

Tools and techniques

Typical graphical techniques used in EDA are:

- Histogram
- Scatter plot
- Heat map

- Bar chart
- Line Polar
- Bar Polar
- Choropleth
- Scatter matrix
- Line plot
- Stack plot
- Area Plot
- Rel plot
- Interactive versions of these plots

Dimensionality reduction:

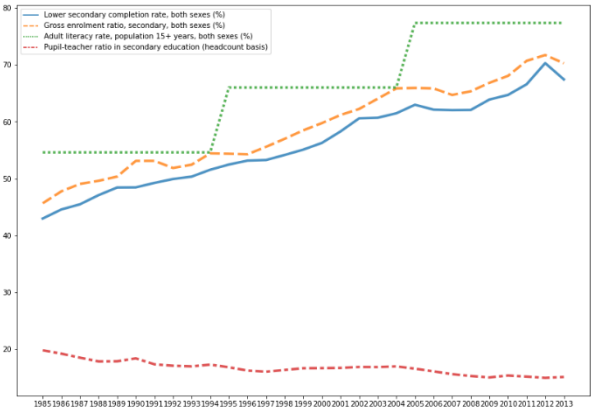
- Multidimensional scaling
- Principal component analysis (PCA)
- Iconography of correlations

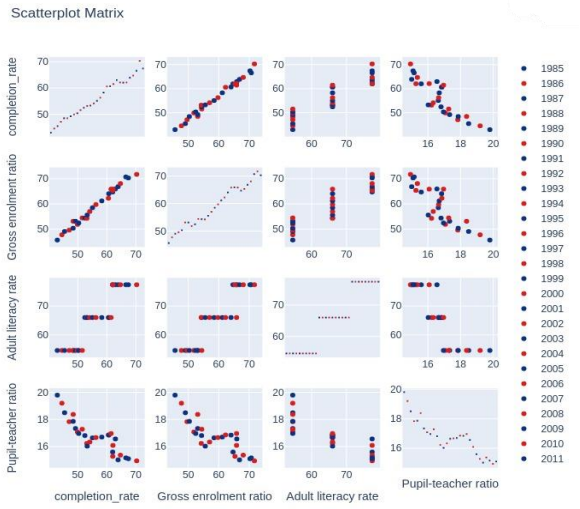
Data Visualization:

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

Problem Statements and Outcomes:

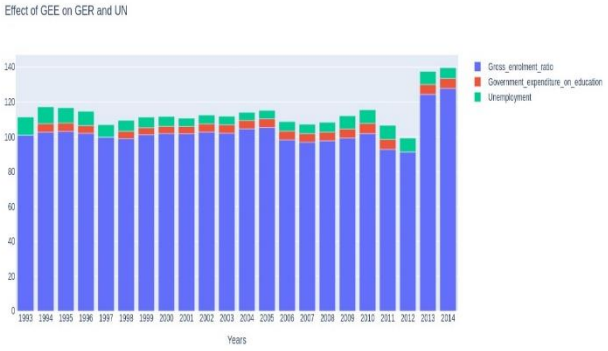
1. Analyse the trends for Gross enrollment ratio(GER), Lower secondary completion rate(LSC) , Adult literacy rate(ALR) and Pupil-Teacher ratio(PTR) for Arab World group of countries. And Compare them with one another.

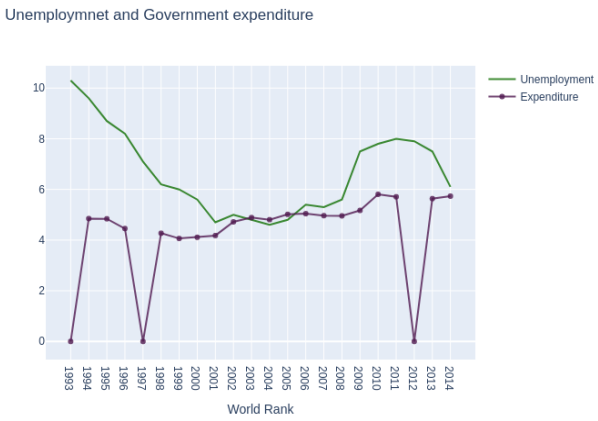




- Literacy rate is calculated in every ten years.
- Pupil-Teacher ratio is decreasing at increasing trends of gross enrollment rate and lower secondary completion rate.
- Increasing trends of enrollment and completion rate is pushing the adult-literacy rate.

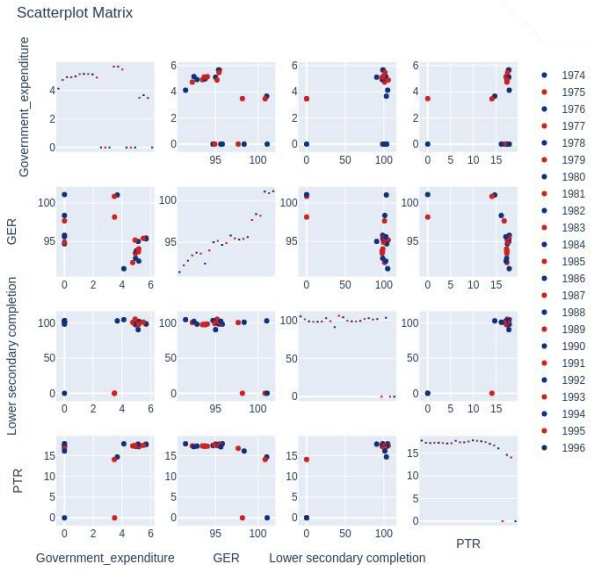
2. Analyse the effects of Government expenditure(GEE) on other education heads for United Kingdom.





- Government Expenditure on Education is following the trends of Unemployment.
- United Kingdom is a developed nation and shows the obvious trends according to the GEE

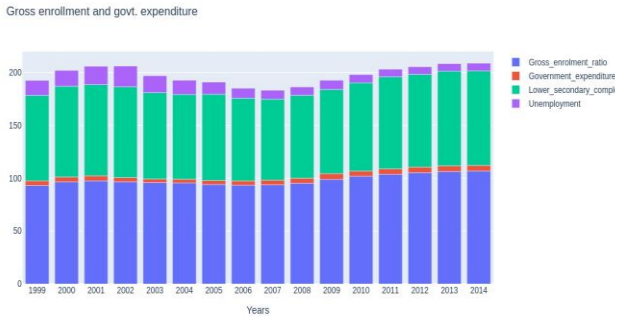
3. Problem Statement 3:Compare the trends of GER,LSC,PTR and GEE for Japan.



- The trends of Japan clearly demonstrate the increasing government expenditure broughts down the pupil-teacher ratio and increase the Gross enrollment ratio and lower secondary completion rate.
- But the chart of Japan also gives one more important insight. When they reduce the government expenditure, the probable outcome should be negative trends in other indicators, but the situation remains as it is.
- Here Japan would have arrived at the most reliable and appropriate percentage of GDP required for their education sector. This is the necessary monitoring of socio economic conditions of any particular country.

.

4. How Unemployment and GEE is closely interlinked in Argentina.

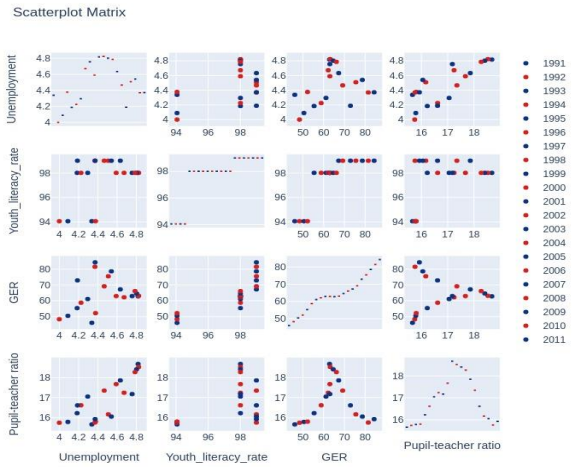


- In 2002 When expenditure lowers to 4 % of GDP, unemployment increases to near about 20 %.
- But later on when Argentina have increased the expenditure in 2014 to 5.3 % of GDP, the unemployment lowers beautifully to 7.3 %.
- Nevertheless 7 % unemployment is still high, according to the standard global requirements.

5.Find out, whether increase in GEE would help Botswana to create positive impact on Employment trends.

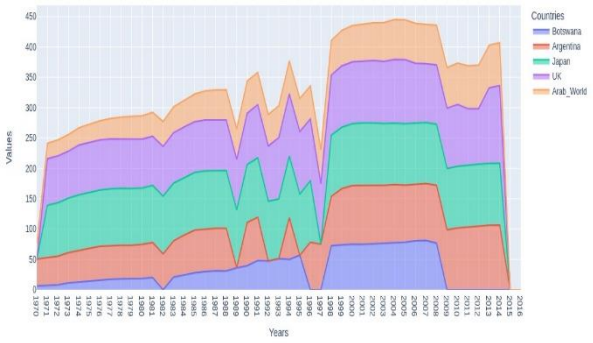
- It has found that Botswana is contradicting the postulates which have been given by earlier studies.
- In Botswana , despite increasing the government expenditure , the unemployment rate is still very high.
- Botswana spend nearly 10 % of GDP on education but failed in improving the employment rate.

6.What is the current status of education in East Asia Pacific nations.



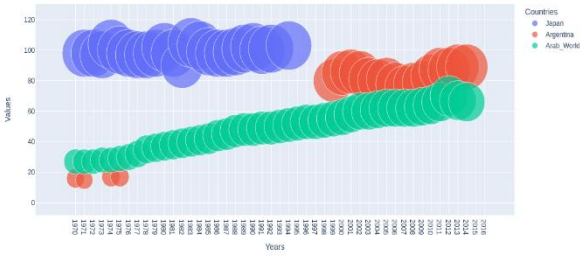
- East Asia Pacific shows all the trend patterns and growth indicators very clearly in conventional manner .
- GER and ALR shows straight forward improvement, uplifting the quality of employment by decreasing the PTR and unemployment.
- Unemployment and PTR shows great coincidence with each other. In the times of 2002 the Unemployment and PTR were at the peak.
- Later on due some probable good policies, both the parameter have reduced to normalcy and touched the Global standards.

7.Compare the GER of all the above mentioned countries and compare the trends.



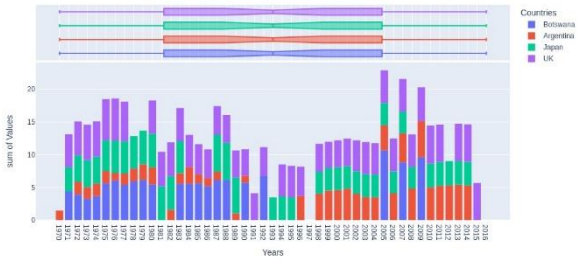
- It can be observed that all the countries are in upward trends in GER, and can be analysed on other grounds.
- Every country is showing the presence of resistance at the same time.
- It might be due to the international disorder.

8.Analyse and compare LSC of all the above mentioned countries and draw the trend fluctuation.



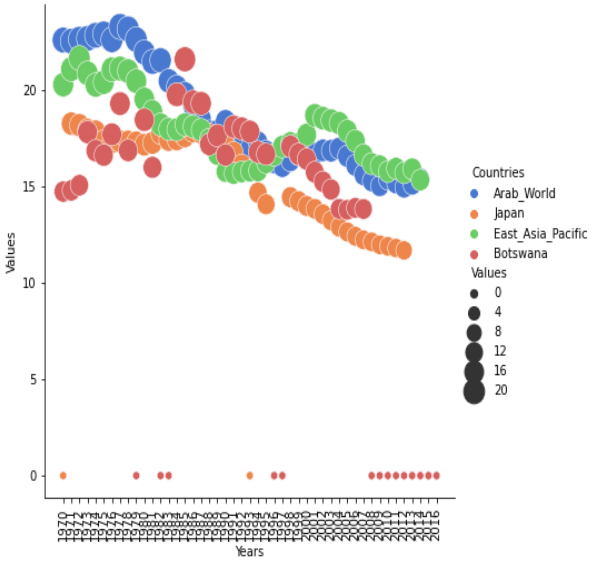
- As mentioned earlier the Arab World group of countries have shown a significant growth in LSC in comparison with other countries.
- Nevertheless it is still below the Global standards.

9.Observe the GEE of all the above mentioned countries and compare it with their present educational status.



- Earlier we have mentioned the expenditure trends of Botswana , which is near about 10 % of GDP in 2005.
- But the other countries are at the little less levels.
- Japan has low GEE , but has balanced unemployment rate.
- Argentina shows the increase in GEE form 1996.This change also could be observed in its unemployment rate

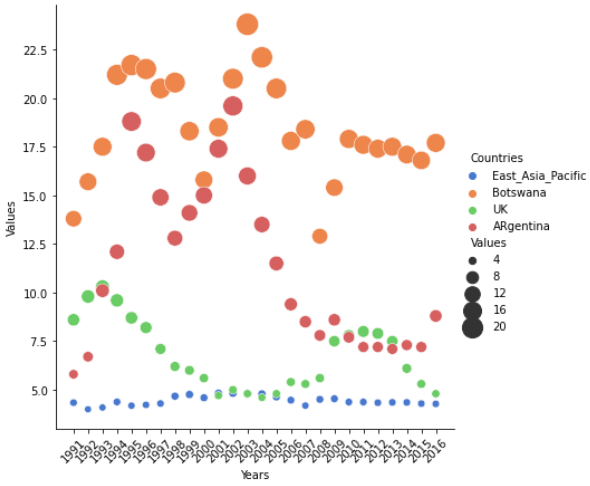
10.Analyse the behaviour of GER, GEE, Unemployment with respect to PTR.



- Every country is showing the downward trends for PTR.

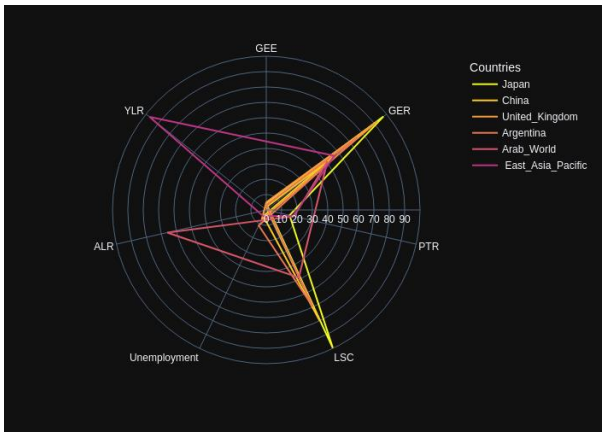
- Arab World and Botswana have shown a significant come back since 1970 to 2016.
- When this map combines with the earlier one, the PTR for Botswana shows the inverse relation with GEE.
- When the expenditure was reduced in 1983 , the PTR was at its peak.

11.Observe the trends of Unemployment, where GEE is low



- Botswana in addition with Argentina showing high levels of unemployment. Even with the passage of time Argentina is reducing with its severity but Botswana is still on a comparative high position.
- When this unemployment trends get compared with the trends of govt expenditure, a close coincidence can be observed. Both are flowing parallely . In next graph both this trends can be observed regarding Argentina .
- UK and EAP shows the coincidence with the standard global trends.

12.Compare all the indicators of every country in one frame, and analyse the comparative movement of indicators.



- The universal truth come in front of us via this line polar chart. It shows those countries with enough government expenditure on education shows the maximum positive trends.
- for e.g Japan with near about 4 % of expenditure shows the positive trends for all the indicators.
- With this analysis , countries those are underperforming could understand, which country they need to study.

Conclusion/Recommendations

1. Despite the positive results , the Gross enrollment and Lower secondary completion rate is not up to the global average for Arab World Countries. The countries should be focusing on, increasing the participation of students and their completion.

2. The direct outcome of United Kingdom chart is , government expenditure on education is very much influencing factor for the changes in unemployment rate of countries.
3. Japan have balanced the public expenditure at 4 % , but still succeeded in making significant enhancements. Countries should follow the case studies of Japan who has the stagnated education indicators.
4. Argentina shows very close connections with public expenditure. In second level of analysis Argentina shows the depressed trends with less expenditure.
5. Botswana spend nearly 10 % of GDP on education but failed in improving the employment rate and pupil teacher ratio.