

SUMMARY

We have been provided a dataset of the World Bank which has 5 datasets EdStats Data, EdStats Country Data, EdStats Series Data, EdStats Country-Series Data, EdStats FootNotes.

Our primary goal is to understand all EdStats Datasets and use our data analysis techniques. Our secondary goal is to draw out actionable insights from our analysis and give conclusions.

As a first step, we saw the dataset and divided the complete project into subcategories that is Cleaning the dataset, Handling the null values, Handling the Outliers, Univariates, and Bivariate.

We started with importing some of the useful python libraries like pandas, matplotlib, and seaborn needed for the project. After the basic operations of importing the dataset, seeing at the overview of the dataset we started with cleaning the dataset. In cleaning the dataset, we mostly focused on the null values, duplicates, and dropping unnecessary columns.

With the help of inbuilt seaborn and matplotlib libraries, replaced the outliers with some mean values and plot various graphs include histogram, scatterplot, boxplot, heatmap, etc., and got some of the following conclusions.

We analyze our datasets in many different ways and we found that some countries have a high percentage of data 10-15% in China, the United States, India, Nepal, Arab world, some have the least data 0-5% in Zimbabwe, Denmark and few counties have average data 5-10% in Algeria, Israel, Brazil, Israel, Russian federation that means least data countries is low literacy rate whereas high data country is high in literacy rate.

Some country has fewer education data that need to improve their literacy rate and development by taking a loan from the World Bank.

Some country's indicators have high data; they can manage and improve their enrollment procedures for the rest of all.

Highly populated countries need to stop their population by giving education to every age group.

Female education is important as male education. Some countries have less literacy rate of females; these countries need to focus on female education

Contributor Roles:

1. Tanjul Gohar

- Data wrangling
- Handling outliers with mean
- Bivariate check
 - a. Analyzed all year's data
 - b. Area-wise population analysis
- Univariate check
 - a. Each Country Analysis
 - India
 - China
 - The United States
 - b. Mean of Indicators
 - BAR.NOED.1519.FE.ZS for India, Nepal, and Russia Federation
- **Conclusion**

2. Kanika Singh

- Data Wrangling
- Handling outliers by mean
- Bivariate check
 - a. Analysis of one Indicator for 7 countries
 - Duration of compulsory education
 - Adjusted net enrolment rate, primary, both sexes (96)
 - Percentage of female population age 15-19 with no education**(India, Israel, United States, Russian Federation, Denmark, China, and World)**
 - b. Country's Name
 - c. Country's percentages
- Univariate Check
 - a. Country of Data percentages
 - India
 - The United States
 - Arab World
 - b. Median Value analysis
 - India and Zimbabwe
 - Bahrain and Belize
 - Belgium and Algeria
- **Conclusion**

All the predictions were made as a team after proper understanding and discussion of each topic.

GitHub Link: <https://github.com/Tanjul5/World-Bank-Education-EDA-capston-1/tree/6f9948ddc004a5e939f2eac66715340a6a3df23a>

Drive Link:

<https://drive.google.com/file/d/1VkhRuXdnmcUDHUOOMrk5u92yXv60huei/view?usp=sharing>