

Capstone Project Submission

Team Member's Name, Email and Contribution:

1. Neel Naik(neelnaik29.nn@gmail.com)
 - Data Wrangling
 - EdStatsData, EdStatsFootNote
 - Loading and Preprocessing
 - Structuring data
 - Enriching data
 - Data Validation
 - Data Mining
 - Row-wise Matrix Collection Method
 - Data Analysis
 - Early Childhood Enrolment
 - Primary Education, Secondary Education and Tertiary Education
 - GDP Expenditure
 - Mean Performance over Mathematics scale
 - Visualizations
 - Bar graphs and Pie charts
 - Debugging
 - Observations
 - Summarization
 - Conclusions
2. Piyush Nirwan (piyushnirwan02@gmail.com)
 - Data wrangling
 - EdStatsData, EdStatsCountry
 - Data Cleaning
 - Summarizing data
 - Enriching Data
 - Data Mining
 - Column-wise Matrix Collection Method
 - Data Analysis
 - Early Childhood Education
 - Government Expenditure on Education
 - Secondary Education
 - Learning Outcomes - PISA
 - Visualizations
 - Trend Lines and Subplot objects
 - Segmentation
 - Presentation
 - Observations
 - Conclusions
3. Prasad Khadatkar (prasadkhatkar18@gmail.com)
 - Data wrangling
 - EdStatsData, EdStatsSeries
 - Loading and Discovering Data
 - Data Classification
 - Data Validation
 - Code Evaluation
 - Data Analysis
 - Female Enrolment
 - Adult Illiterate Population
 - Tertiary Education
 - GDP of countries

- Visualizations
 - Plot and Horizontal Bars
- Debugging
- Summarization
- Presentation
- Observations
- Conclusions

4. Saransh Jain (jain.saransh97@gmail.com)

- Data wrangling
 - EdStatsData, EdStatsCountry-Series
 - Loading and Discovering data
- Data Analysis
- Data Cleaning
 - Removing Null
 - Removing Duplicates
- Data Visualization
 - Pie chart
 - Donut chart
 - Line Chart
 - Bar chart
- Segmentation
- Observations
- Conclusions

5. Shivank Shukla (shiva999111222@gmail.com)

- Data wrangling
 - EdStatsData ,EdStatsCountry
 - Loading and Discovering data
- Data Analysis
- Data Cleaning
 - Removing Null
 - Removing Duplicates
- Data Visualization
 - World map
 - Plot
 - Presentation
- Segmentation
- Presentation
- Observations
- Conclusions

Please paste the GitHub Repo link.

Github Link:-

<https://github.com/NeelNaik29/EDA-Word-Bank-Global-Education-Data>

Drive Link:-

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

The World Bank Group is one of the world's largest sources of funding and knowledge for developing countries. Its five institutions share a commitment to reducing poverty, increasing shared prosperity, and promoting sustainable development. The World Bank Ed-Stats Query holds around 2,500 internationally comparable education indicators for access, progression, completion, literacy, teachers, population, and expenditures. The indicators cover the education cycle from pre-primary to tertiary education. The query also holds learning outcome data from international learning assessments (PISA, TIMSS, etc.).

As a first step we explored through all the files in the dataset to understand the rubrics and content of dataset that we have. The EdStatsData seemed to be the most important file as it holds all the Indicators data for countries associated with world bank. Now we selected a few indicators which seemed relevant for our analysis and generalized them based on 5 categories, they're Early Childhood Education, Expenditures, Literacy, Population and Learning Outcomes.

Now we created two functions that mines out a small chunk of indicator data for selected countries out of the data file in a row-wise or column-wise manner. Row-wise dataframe is a clip out of the data such that countries are indexes of dataframe and columns represents the year values, vice-versa in column-wise dataframe. With these functions we were able to extract the meaningful data for various indicators, also we cleaned the data by dropping unnecessary rows or columns and by filling null values with mean and Backward/Forward fill methods.

For **Economy**, we have found that most countries with high per capita income are in Europe and in north America with per capita more than 4 times that of the world average. On the flip side most lower income countries are in Africa with per capita income below 2000USD which is unsustainable.

For **Early Childhood Education**, we found that for Enrolment in early Childhood education, the uptrend is significantly positive. It is evident that India is making remarkable and positive growth in this field. The early childhood enrolment rate of India shows tremendous extension, which signifies that people are more conscious towards the education of early-childhood education.

For **Literacy**, after the year 2006, we see that the illiterate population metric has been decreasing significantly in India; we could relate that with the increase in the government expenditures and initiatives towards education. Also the literacy rate for India has shown a whopping 30 percent hike over the span of 30 years (1981 to 2011).

For **Population**, We observe a significant leap in the population of official age for pre-primary education in India, during the period 1997-1998. We are also able to see that across all the graphs, India is on a consistent incline. For the countries like Japan, United States and Germany, we see an overall stable trend with minor fluctuations. This shows a potential focus towards stability and consistency in the population.

For **Expenditures**, we are able to derive that the total expenditure of the government on education for India and Sri Lanka decreased in 2014, than that of 1999, and the expenditure by the United States and the United Kingdom has consistently increased over the years.

For **Learning Outcomes**, Among the Asian countries, China has been holding the lead in PISA scores for Reading and also for Mathematics; Finland leads amongst European countries.

From the analysis we are able to conclude that India among all is one of the most significantly developing countries. India has shown positive growth over the period of last 20 years (considering from 1995 to 2015). The early childhood enrolment rate of India shows tremendous extension, which signifies that people are more conscious towards the education and also the literacy rate has improved considerably. Similar trends for countries like USA. USA is currently more prominent and developed in terms of education but the outstanding growth of India over the years shows great potential for the Education domain.

