School of Information Sciences

(A Constituent Institute of Manipal University)

Data Analysis of Literacy and Level of Education

Mohit Awana 171046031

Course: ME- Big Data and Data Analytics

Under the Guidance: Mr. Balaji B

INDEX

ABSTRACT

Literacy is a key skill and a measure of a population's education. Literacy rate have risen drastically, only 12% of the people in the world could read and write in 1820, today only around 20% of the world population remains illiterate. The objective is to find out literacy rate across the globe, how much the factors (attendance rate, enrolment rate) affect the literacy rate and how much literacy differs according to gender.

INTRODUCTION

Literacy is traditionally understood as the ability to read, write, and use arithmetic. The modern terms include the ability to use language, numbers, images, computers, and other basic means to understand, communicate, gain useful knowledge.

UNESCO defines literacy as the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts.

Literacy is an important skill when it comes to personal growth, culture and development. Nowadays, it has become even more important because of the way our society and economy are shaping.

Challenges with literacy affect the people around the globe. These challenges have impacts on workforce, economy and day to day lives of countries.

SPECIFICATIONS

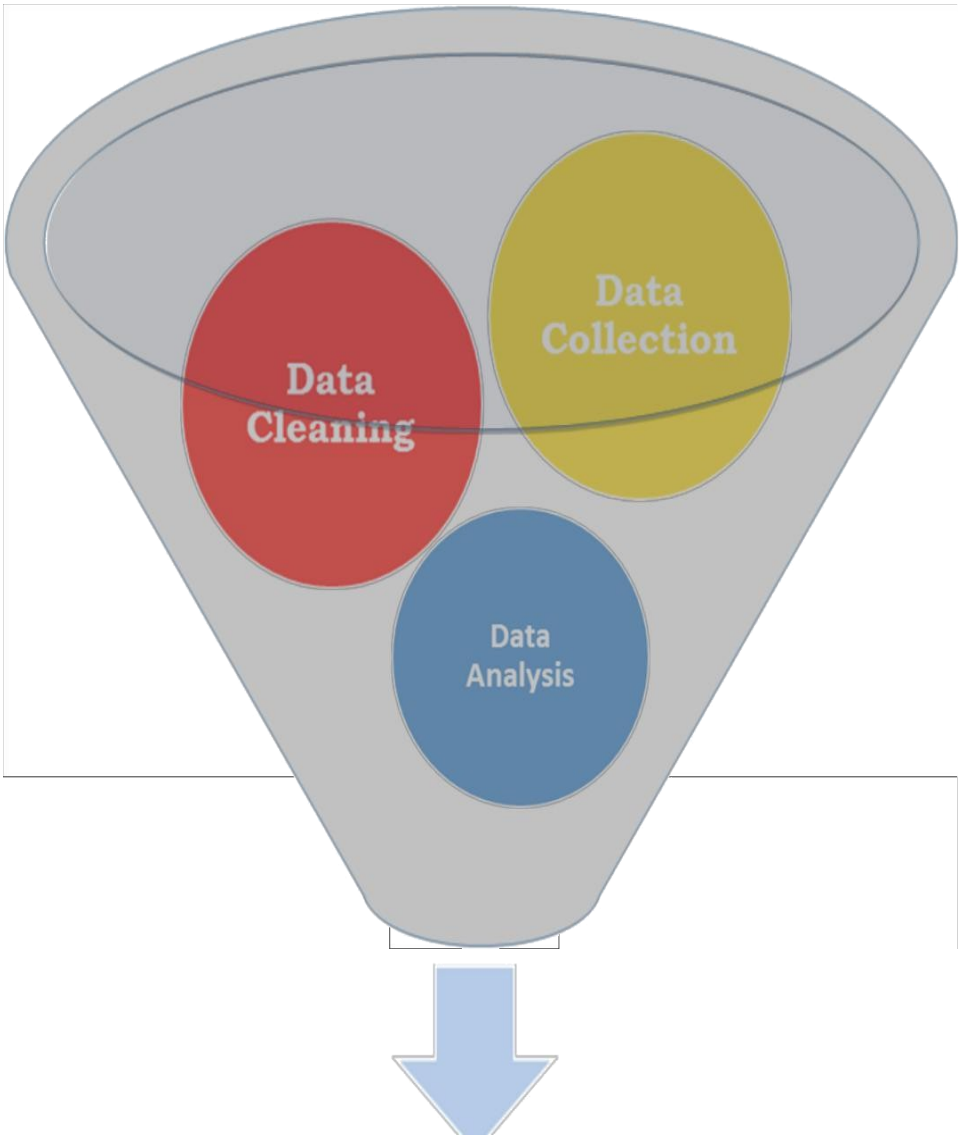
Software Requirement:

OPERATING SYSTEM: Windows, Ubuntu, OS X
PROGRAMMING LANGUAGE: Python 2.7 or above
PYTHON PACKAGES: Numpy, Pandas, Matplotlib, NVD3

Hardware Requirement:

PROCESSOR: 64 Bit Intel or AMD
PRIMARY MEMORY: 4 GB or above
SECONDARY MEMORY: 50 GB or above

METHODOLOGY



Data Collection

Data extraction is the process of retrieving data out of data sources for further data processing or data storage. Data extraction is applied when experimental data is first imported into a computer from primary sources.
Extracted datasets from data.unicef.org (UNICEF)

Data Cleansing

Data cleansing refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data. Incorrect data leads to false conclusions. Cleaning of the datasets was done using **Pandas** and **Numpy (**Python library).

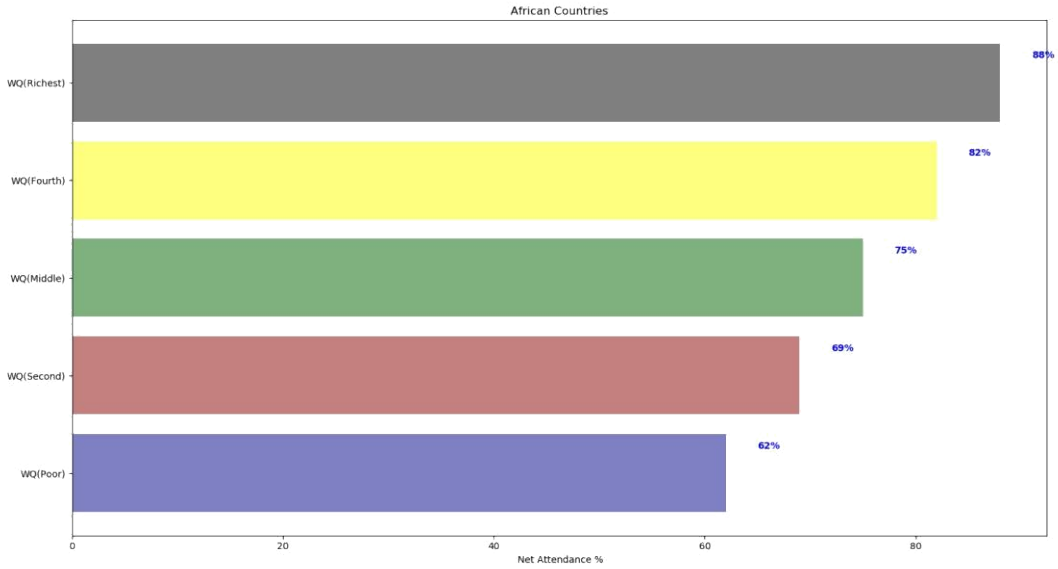
Data Analysis

Data analysis: is inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information. Pandas stands for Python Data Analysis Library. It's a powerful tool for data analysis. Helps you clean up and put data together into a format that is easy to use, excel friendly, and analyze. Countries were divided according to the continents and analysis was done on respective countries on the basis of wealth quintiles, net attendance rate, enrolment rate and gender wise.

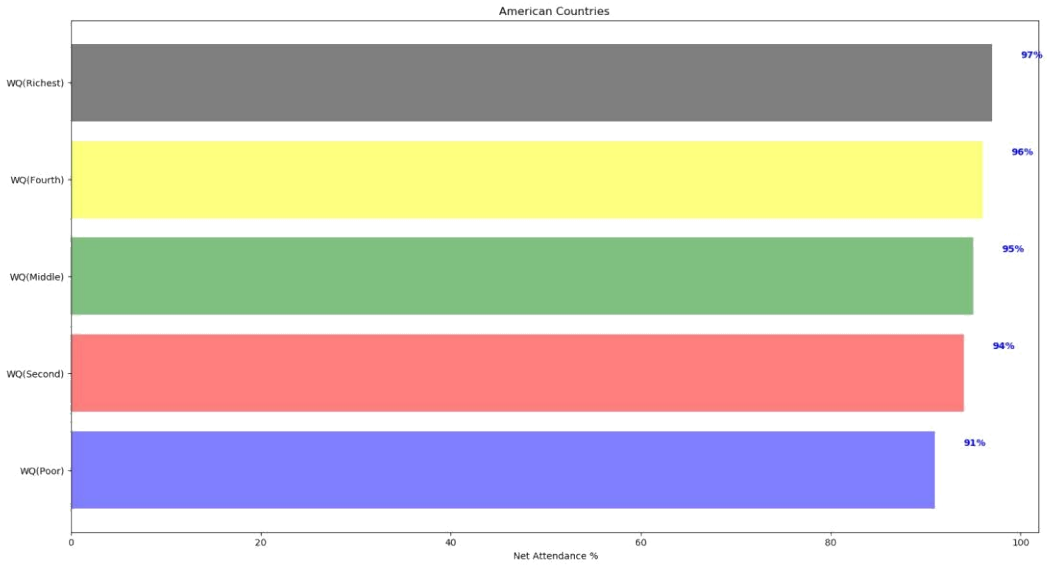
Data visualization

Data visualization helps people understand the significance of data by placing it in a visual context. Visualization is central to advanced analytics. Matplotlib, NVD3, plotly were used to visualize the data into meaningful context .Plotted various graphs according to literacy rate, net attendance rate, net enrolment rate.

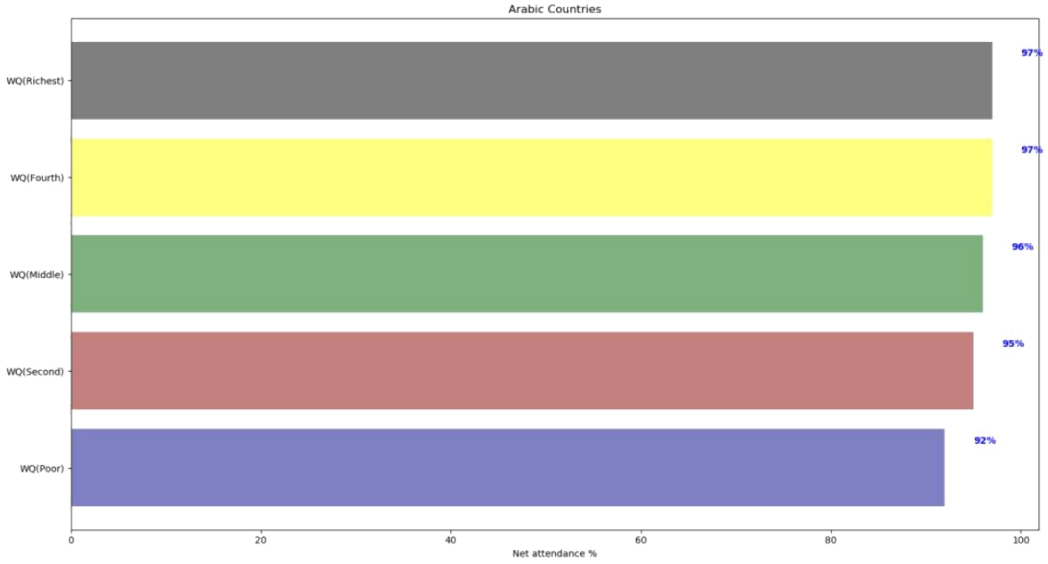
RESULTS



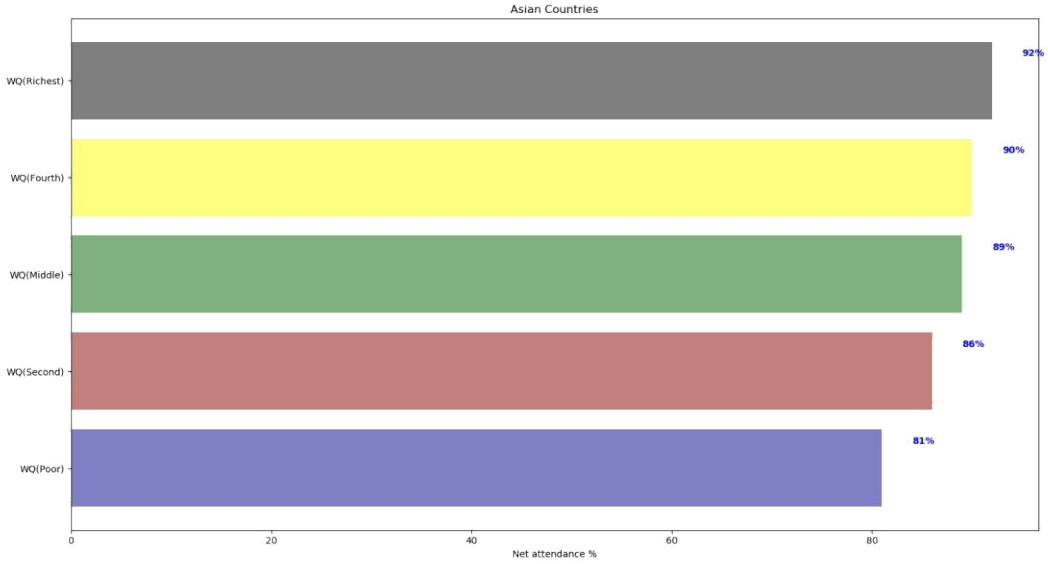
1. The graph of Wealth Quartile over Net Attendance rate for the African Countries



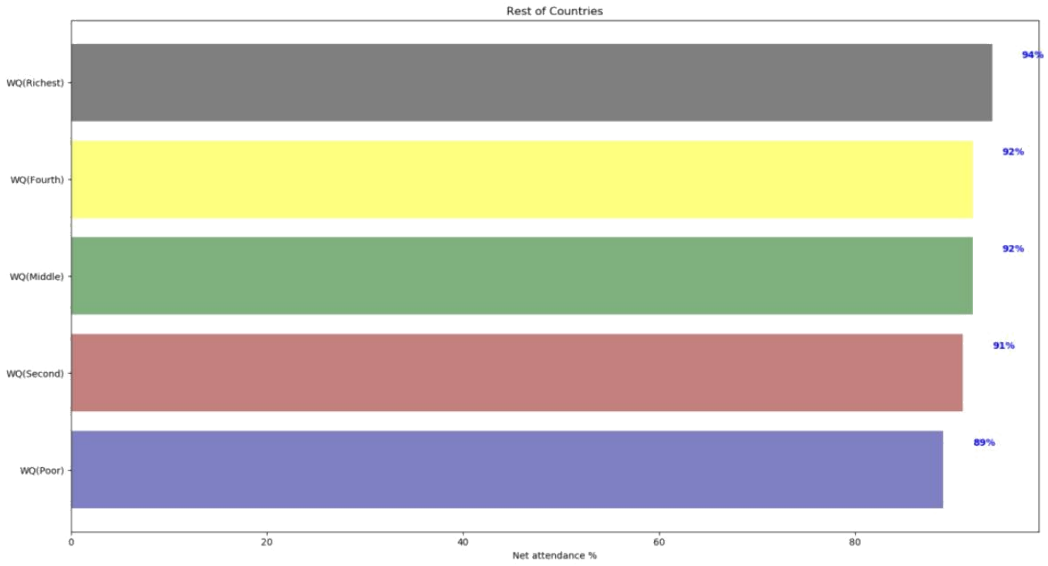
2. The graph of Wealth Quartile over Net Attendance rate for the American Countries



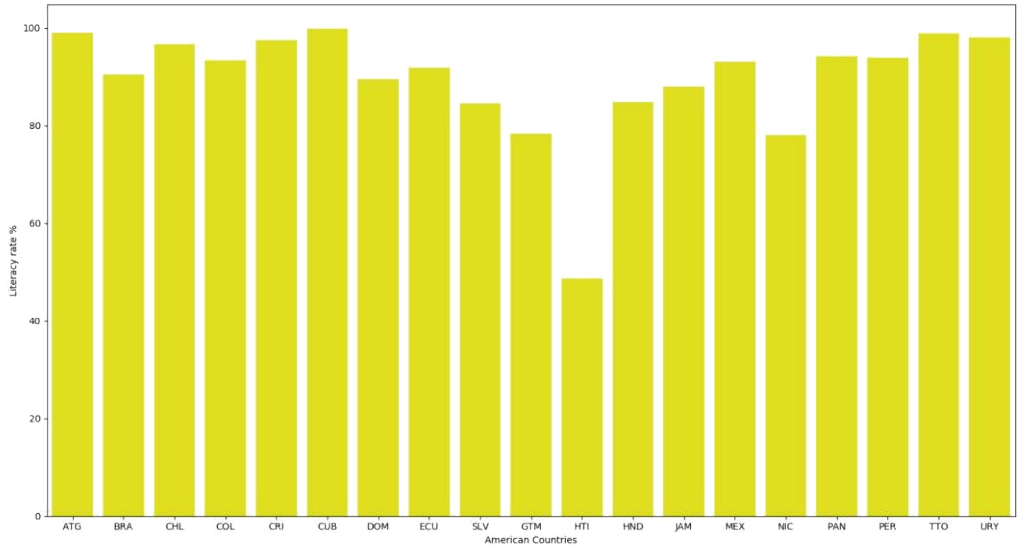
3. The graph of Wealth Quartile over Net Attendance rate for the Arab Countries



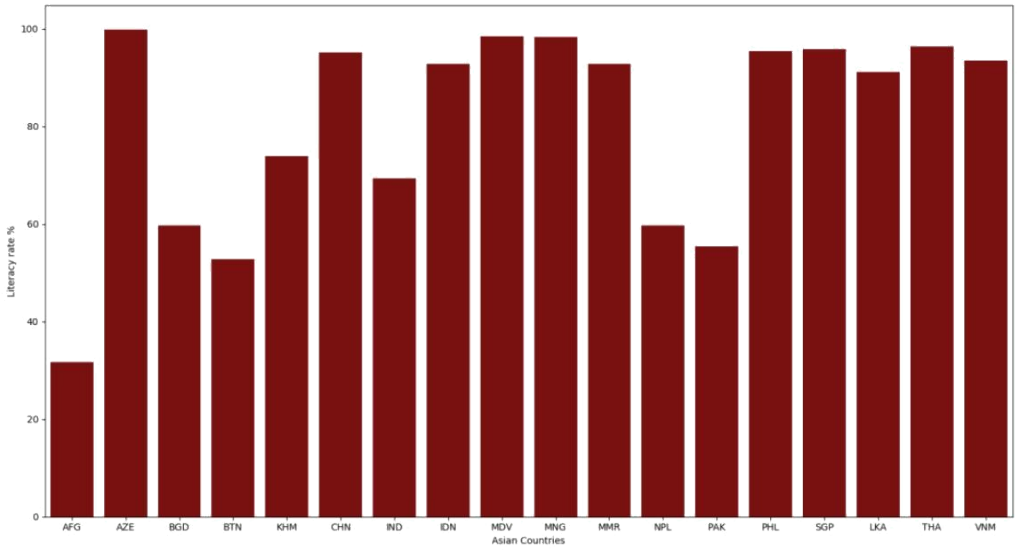
4. The graph of Wealth Quentile over Net Attendance rate for the Asian Countries



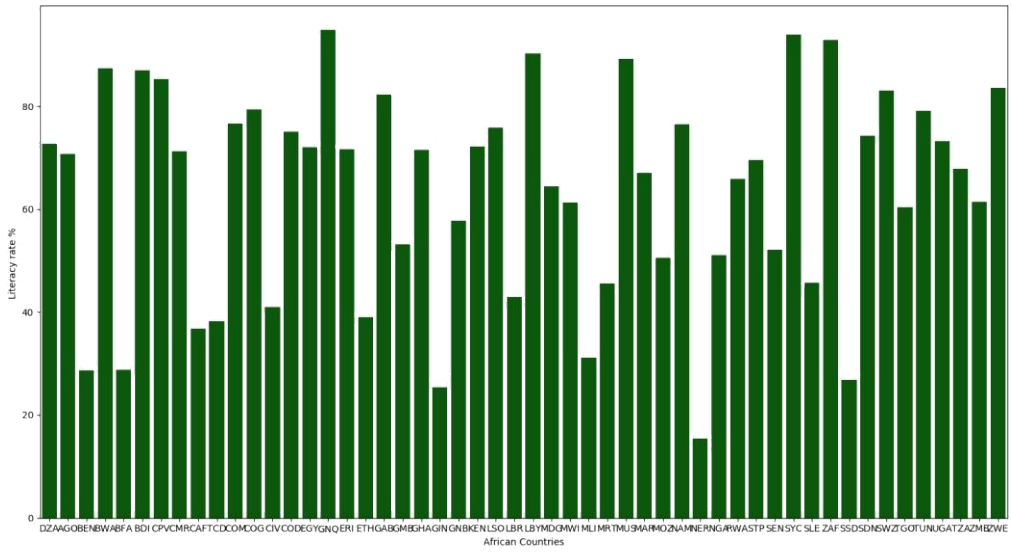
5. The graph of Wealth Quartile over Net Attendance rate for the rest of the Countries



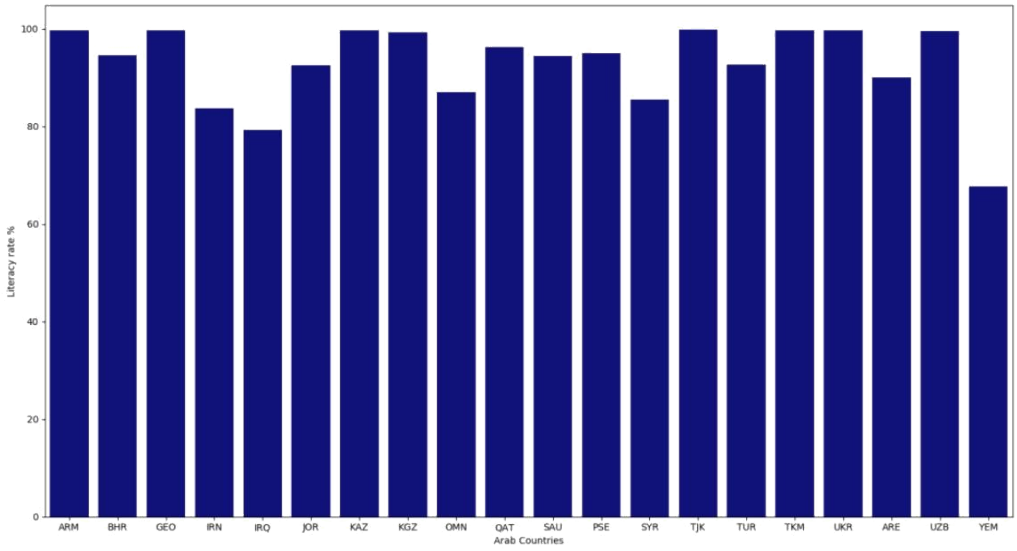
6. The graph indicating the levels of literacy rates in American Countries



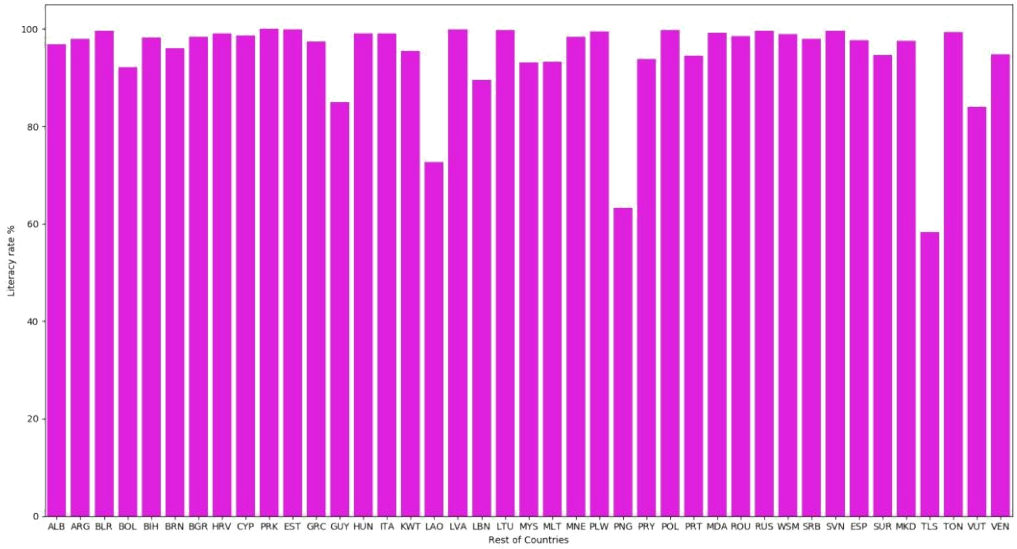
7. The graph indicating the levels of literacy rates in Asian Countries



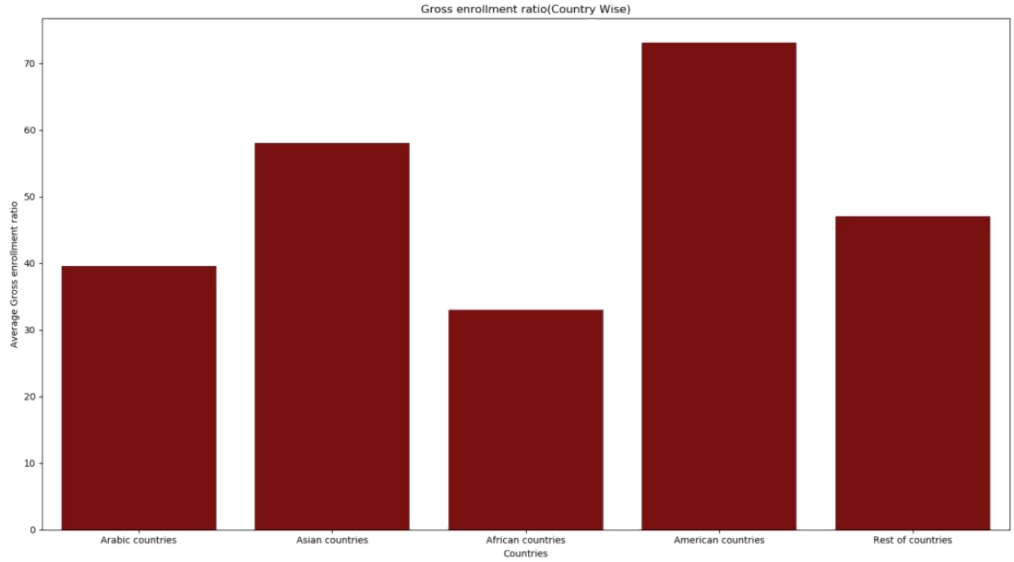
8. The graph indicating the levels of literacy rates in African Countries



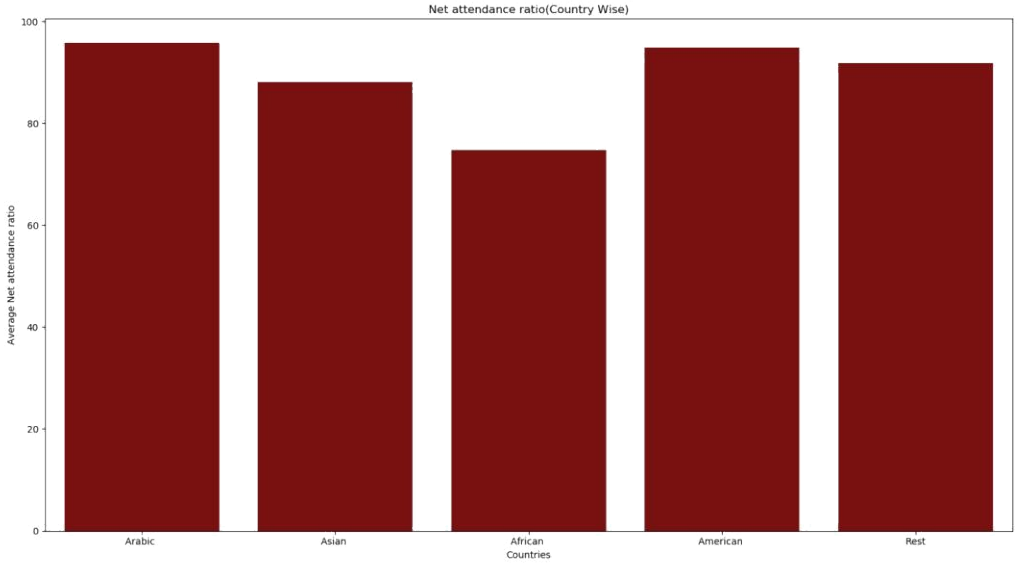
9. The graph indicating the levels of literacy rates in Arab Countries



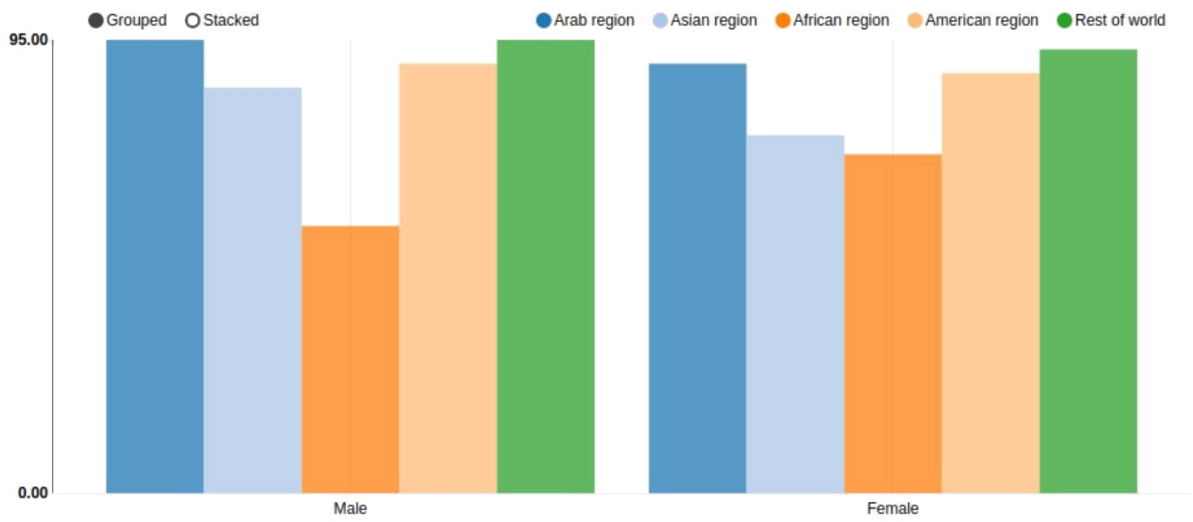
10. The graph indicating the levels of literacy rates in the rest of the Countries



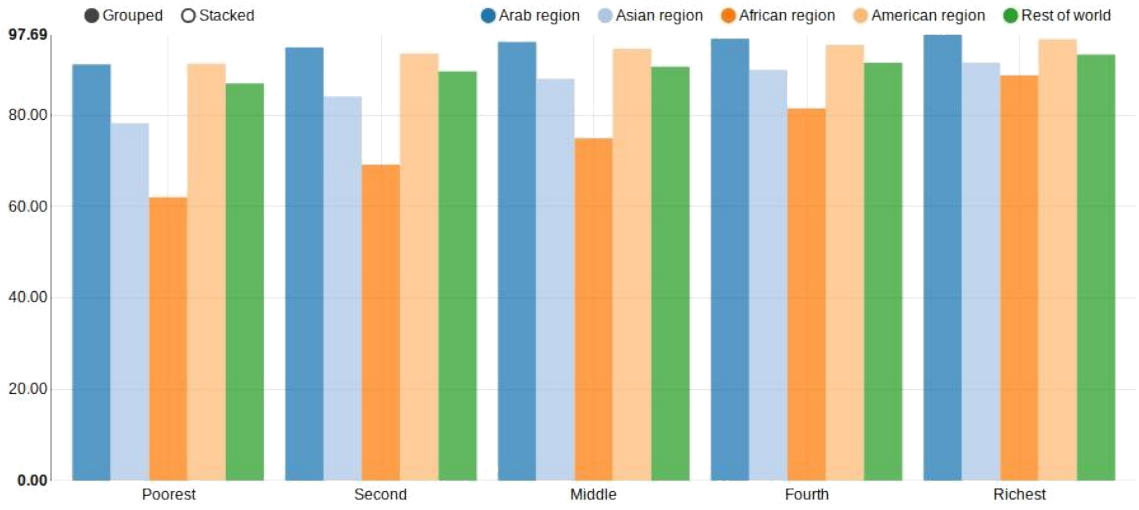
11. The graph indicating net enrollment ratio according to Countries.



12. The graph indicating net attendance ratio according to Countries.



13. The graph indicating literacy rate among males and females according to Countries.



14. The graph indicating literacy rate over Wealth Quartile distribution according to Countries.

SCOPE OF FUTURE WORK

The collected data can be used to predict future literacy rate by using machine learning algorithms like multiple regression.

This can have various positive impact on determining self-esteem and empowerment, political participation, democracy and ethnic equality cultural change and preservation of cultural diversity, health, reproductive behavior, education and gender equality.

INDIVIDUAL CONTRIBUTION

I collected the dataset from data.unicef.org. Implemented cleansing technique using pandas to clean the dataset for analyzing the data with accordance to the project. Used various visualizations for depicted the data into meaningful information in form of graphs. Matplotlib , NVD3, Seaborn were used to visualize the data.

BIBLIOGRAPHY

The datasets were extracted from data.unicef.org