

EDUCATION & INNOVATION: USING EDUCATION EXPENDITURE AND OTHER FACTORS OF INNOVATION TO PREDICT GDP PER CAPITA

Team #34 | Chiamaka Anonyuo, Jahmarli Cohen, ShaTori Harris, Stephanie Wilks

Highlights

- As a country's education expenditure GDP percentage increases, GDP per capita also increases
- Further analysis will be needed to create a stronger and more accurate model when innovation is factored in

Background

Two elements that may allude to a country's ability to function are education and innovation. Our theory is that a country that spends more on public education may exhibit more signs of innovation (i.e. patents, internet usage, etc.) and a higher GDP per capita. Through our analysis, we hope to persuade world leaders and prominent stakeholders to consider the effects of public education investment on a country's economic health.

Data

~10.3K
Records

>150
Countries

21
Years

The data was sourced from the UNdata website. The tables included information on public expenditure on education, GDP, GDP per capita, and innovation factors for over 150 countries from 2000 to 2021.



Data Processing

32
Observations

8
Countries

4
Years

11
Features

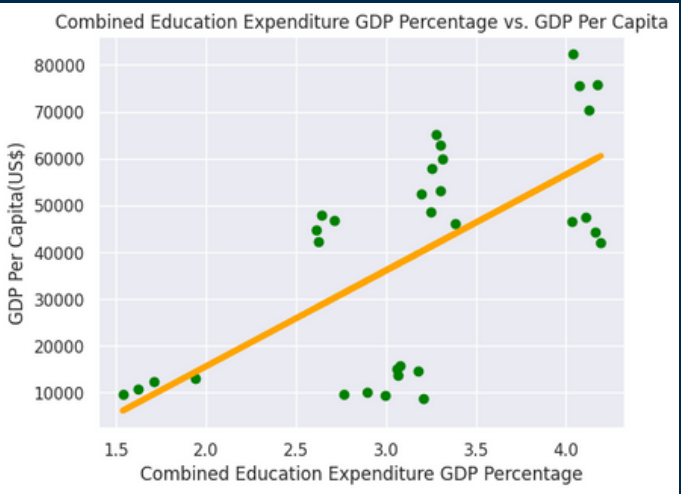
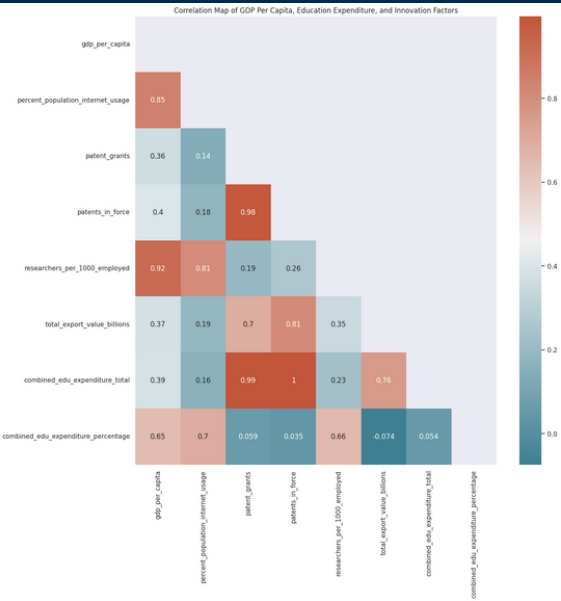
The raw data contained in the seven csv files pulled from the UNdata website were processed in a Jupyter notebook. The Pandas library was used for cleaning and preparing the data for exploratory data analysis. The Matplotlib and Seaborn packages were used to explore and chart the data. The Statsmodel package was used to create a Multiple Linear Regression model with our sample data. Finally, we used Tableau Public to create a dashboard for stakeholders to view. That dashboard can be viewed using the following link: <https://tinyurl.com/ybwema5n>

Conclusion and Next Steps

We were unable to reject the null hypothesis for this analysis, but we believe that we can create a stronger model that will meet the statistical significance threshold.

As for next steps, we will include more countries in our sample data and will use the PloynomialFeature library in the sklearn package to perform a polynomial regression.

Model



<0.05
p-score

0.42
R-squared

We started off with a simple linear regression model where GDP per capita was the dependent variable and combined education expenditure percentage was the independent variable. This model explained 42% of the variability in our dependent variable and was statistically significant with a p-score less than 0.05.

Next, we performed a multiple linear regression with our factors of innovation and encoded dummy variables that stood in for the eight countries of our sample data. After several iterations, we created a model that had an Adjusted R-squared value of 0.993 and a Condition number of 192. However, none of the variables in the model were statistically significant.