Abstract

Ecuador has seen a significant improvement in it's economy over the past several decades. Despite economical hardships early in the nation's history- an increase in tourism, job opportunities, and exported goods have allowed for large areas of the population to see a drop in poverty. One of the proposed solutions by the government to tackle the widespread poverty in the country, was to provide free, or inexpensive, primary and secondary schools that were government funded across the country - allowing citizens and children to have an opportunity to receive an education at no cost. While the 1990's saw a big increase in enrollment numbers and economic boosts after the implementation, in recent years those numbers have dropped or plateaued in specific areas of the country- and have increased the poverty gap as coastal cities continue to grow.

Design

The proposed problem for this project is to further investigate the areas of the nation that have higher poverty rates, and analyze how the current methods of the education system may have provided unequal assistance across the population. By using the education systems most recently accessible dataset, providing a model that can pinpoint the areas that continue to struggle under a poverty trap, and how much education (or lack of education) has an impact on each area. By providing an analysis and model, a redistribution of schools can be considered with a larger focus on geography, and perhaps a different sourcing of fundings for areas that may need more assistance than others.

Data

The data used for this project ranged from accessing public data from the World Bank Data, for economic summaries and governmental department spendings across the world - including Ecuador, to gathering provincial data from the Ecuadorian census bureau and Ecuadorian education departments that are publicly accessible. The World Bank dataset provided a clear picture of the status of the nation over a large period of time, however, missed the granular details that were needed to be found on a provincial level. The data bases found were also gathered from Kaggle in CSV and XLSX format.

Additional geographical data was also found for Ecuador's provincial locations - using shapefiles across government databases as well. These were used for the geographical visualizations of poverty distributions across the country.

Algorithms

For the EDA portion of this project, the main spreadsheet functions were used to merge all data tables. VLOOKUP's were used across different datasets to link provinces across tables, leading to a summarized table with all main province attributes. Similarly, aggregation functions and pivot tables were used to calculate counts of schools by province, calculations of averages of population percentages, and grouping of provinces in regions.

Python was used as well to create a simple regression model with the summarized data to find the highest correlations across different features in our dataset. The target variable used was percentage of population under poverty, and used key features relating to schooling such as- institutions per capita, illiteracy rates, and average years of schooling.

Finally, Tableau was also used for some further data exploration, as well as primary tool for selecting visualization options and the display of geographical data for distributions of different features across the country.

Tools

- Google Sheets were primary used for a majority of the EDA performed on all tables, as well as simple data visualizations
- Tableau was used for a majority of the visual displays of data, as well as utilizing the shapefiles that were accessible for geography data of Ecuador
- Light further EDA was used with SQL for broader analysis of spreadsheets
- Python was used for the simple regression model in summarized data