Incidence of Economic Growth and Education Variables in Mortality in the 21st Century

Multivariable Analysis, Predictions and Visualization | Team #6

Carlos Chávez, Yohana Sandoval, Daniel Duran, Elder Giraldo, Fernán Marsiglia, Juan Álzate, Marta Ordoñez, Mateo Moreno, José López.

# Business Problem

Some research has found a positive correlation between income inequality and some indicators tracked by the WHO such as mortality (measured as a mortality ratio). In addition, inverse relationships have been observed between economic development and education in mortality, mainly at early ages.

# Business Impact

Economic growth and education may reflect the effects of other socio-economic variables that may also be related to mortality. For this reason, it is important to conduct comprehensive and interdisciplinary analyses that consider as much information as possible. Such information can be relevant to governmental decision-making on the direction of each country and/or region. This can enable the design and implementation of public policies aligned with the needs of the population and the objectives and goals of sustainable development (Agenda 2030).

Global, regional, and national statistics on population. Also, health and mortality indicators are essential in order to assess development and health progress to guide resource allocation. Specifically, this data is used to monitor progress towards health-related targets within the Sustainable Development Goals (SDGs).

# 

# Data

| **Dataset** | **Source** | **Description** | **Variables** |
| --- | --- | --- | --- |
| **Mortality** | <https://platform.who.int/mortality/themes/theme-details/mdb/noncommunicable-diseases> | **WHO** mortality database contains mortality percentages around the world. | * Age group * Sex * Mortality Rate |
| **Education** | <https://data.worldbank.org/indicator> | **The World Bank** Education database contains some of the most important measures that are part of the direct and indirect measures of education. | * Children out of primary school * Government expenditure on education, total (% of GDP) * Literacy rate, youth total (% of people ages 15-24) * Primary completion rate, total (% of relevant age group) * Unemployment, total (% of total labor force) * Pupil-teacher ratio, primary * Literacy rate, adult total (% of people ages 15 and above) |
| Economic growth | <https://data.worldbank.org/indicator> | **The World Bank** Economic Growth database contains some of the most important measures in the field of economic and economic growth. Most of the variables are related to petrol or energy as an indirect measure of economic growth. | * GDP growth (annual %) * GDP per capita growth (annual %) * Exports of goods and services (% of GDP) * Energy use (kg of oil equivalent per capita) * Fossil fuel energy consumption (% of total) * Renewable energy consumption (% of total final energy consumption) * Revenue, excluding grants (% of GDP) * Inflation, GDP deflator (annual %) |

*It may be considered some other variables that can reflect either Economic Growth or any type of educational index. However, the option of adding a different dataset to those mentioned above is not ruled out.*

# Methods

## Visualizations

# The type of analyses that are going to be carried out lends itself to a dashboard that illustrates the findings of the work. This dashboard should be interactive in order to look at the data from different approaches and in a user-friendly way.

* The Dashboard could have a world map with a gradient for the different numerical variables as well as for the ordinal categorical ones.
* Some checkboxes can be used to filter the categorical data. For example, to filter regions or countries.
* Another important visualization item could be a slicer button, so that users can explore any date throughout the whole 21st century.
* Scatterplots that help users know how the models were fit and plotted ranges that show the prediction intervals are also key parts of the visualization process.
* One section is focused on the context of education, a second section is about mortality and another one for economics with the use of head maps. .
* A final section that relates all three variables together.

To illustrate the data, analytic reporting tools such as Power BI or Tableau will be used. However, for the EDA (Exploratory Data Analysis), Seaborn and Matplotlib would be the most suitable libraries to be used.

## Models

Modeling is a key step in any Data Science project. Here is where the magic happens. For that reason, choosing the correct model or fitting it properly would make the difference in a well-done job or a standard one.

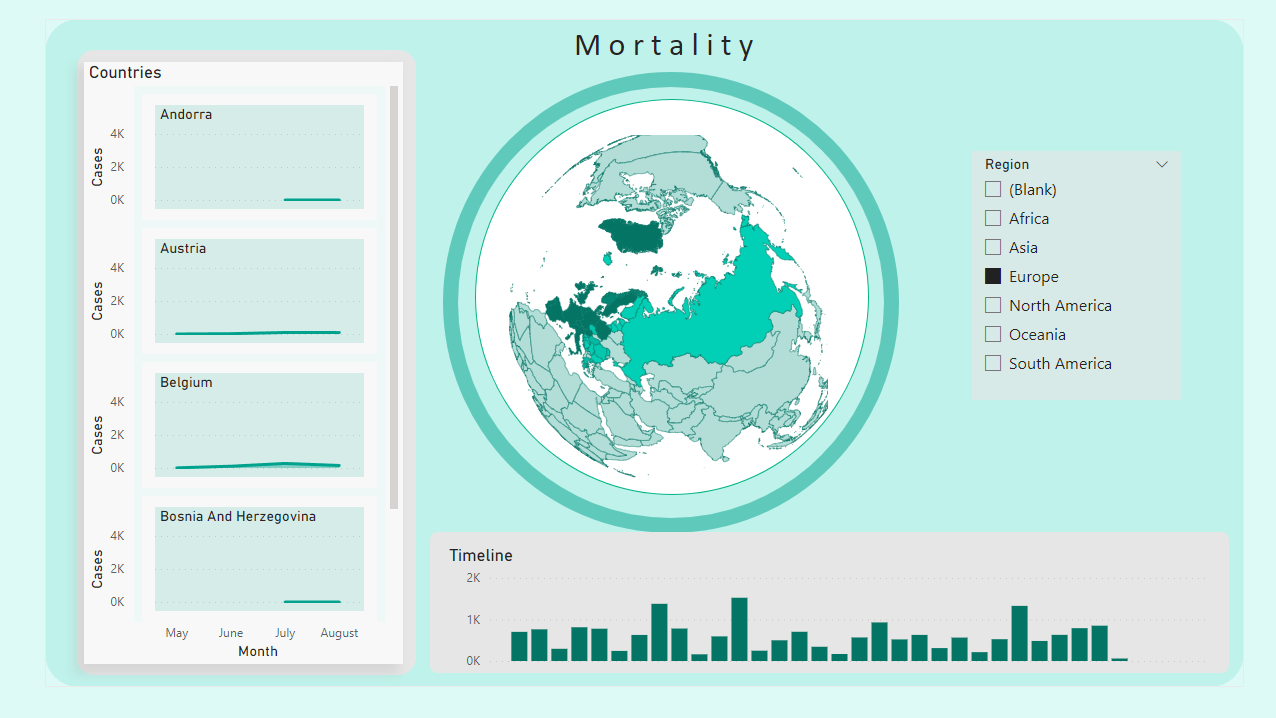
For the supervised model, we know the outcome contained in the mortality table, the economic growth table and the education table. We also know the variable of interest. This variable will allow us to show the incidence and at the same time it can also allow us to categorize the countries by ranking the results obtained.

We believe it may also be possible to work in an unsupervised model where we do not consider a variable of interest; simply with the data, we can perform a clustering and as a result we may find pattern similarities between the data from some countries.

* PCA (Principal Component Analysis): The odds of ending up with quite a few variables is fairly high. For that reason, a dimensionality reduction may be carried out in the process of making predictions about the future. After the reduction, a ML algorithm can be used to predict the outcomes in an unsupervised analysis. However, which algorithm is going to be used is still unknown.

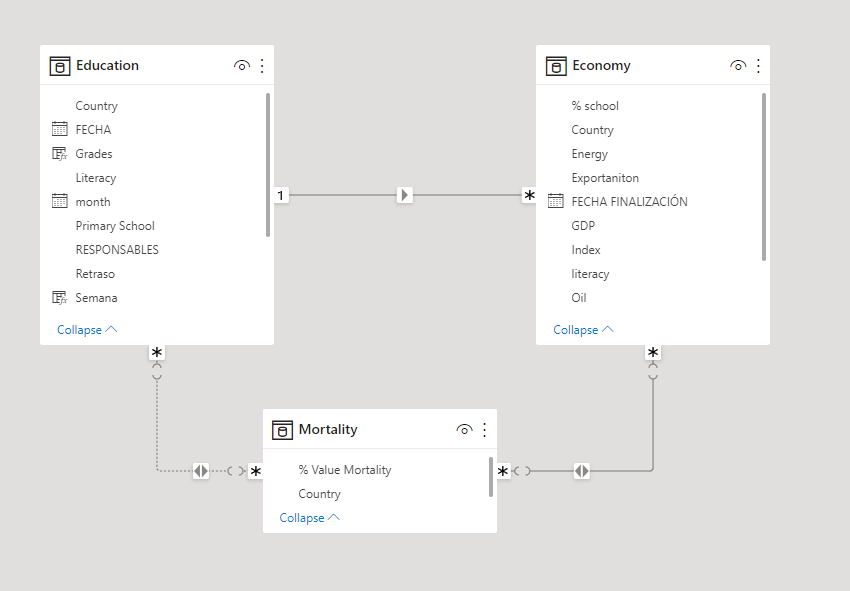
# Interface

An interactive dashboard is a great way of visualizing the data and the insights of it. This should consider regions and countries as well as some line charts and filters that make the interaction with the user useful and user friendly. As it was mentioned before, Microsoft Power BI or Tableau are the options that suit best for this task.

**

The key variable of the databases will be *country*. However, there could be some forein keys such as years or regions. Here is a model that helps understand how the relational (SQL) databases would be interconnected to each other.

***Note:*** *This is just an example.*

**

# Milestones

# 

# 

# 

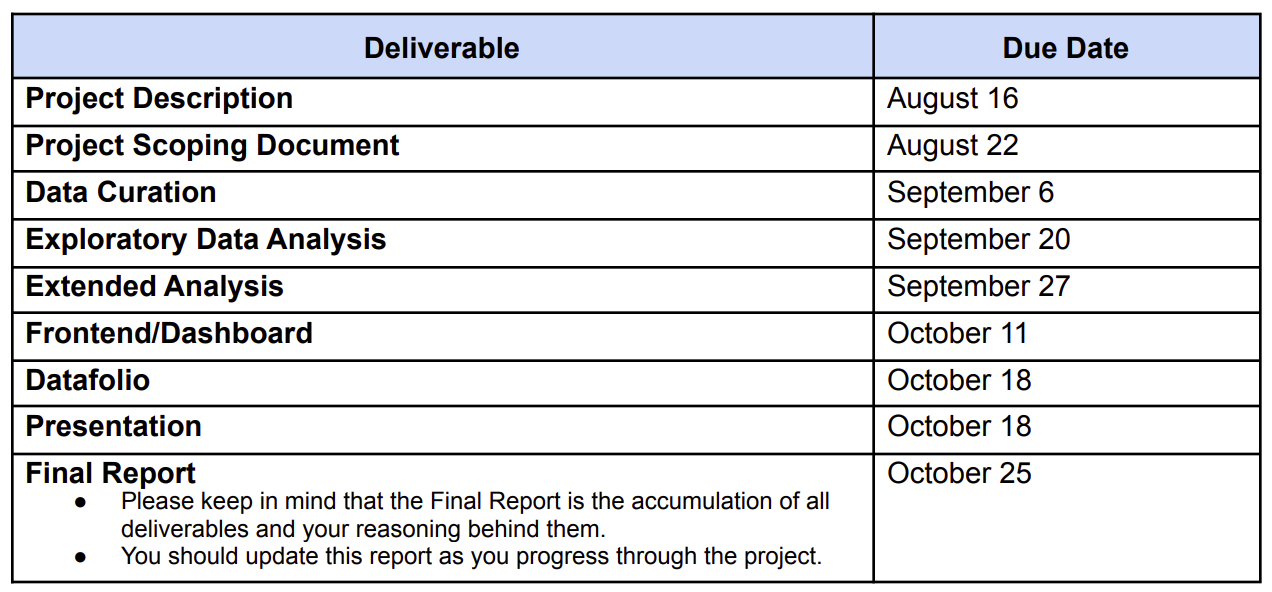
# 

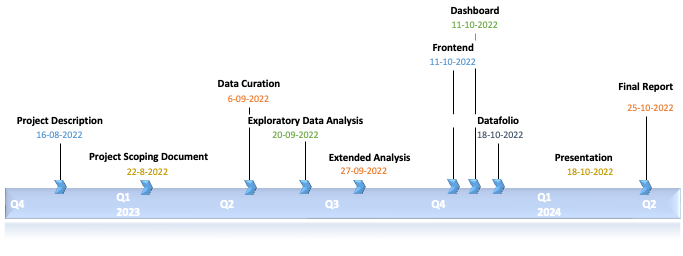
# 

# 

# 

# Timeline





## Concerns

All the analysis is based on the data provided. Any bias in the collection of the data may result in erroneous conclusions.

Some countries have quite a few missing values, and they tend to have missing values in more than one variable. That could lead to an incomplete analysis or can make the project take other objectives.

There are no experts in the group that know exactly all the details of the Data Analysis process. For that reason, there might be other more specific and complex approaches that may not be considered by this team.

Some countries do not report mortality to the WHO. For other countries, the data they provide to the WHO, is not in the standard code or does not have the standard code. For that reason, those values are not shown. In many low-income countries, information about the death cause is difficult to obtain. It is mainly because either the system for recording such information does not work or it does not exist. Additionally, one of the big problems is the lack of medical certifiers to fill up the death certificates.