**Data Preparation Explanation**

1. I am working with three data set as follows
2. World Bank data with 24.178.176 observation.
3. ﻿Quandl API to get real time stock data
4. Cloropeth dataset to create mapping data
5. The World Bank dataset is multidimensional, as follow:

Graphical user interface, text, application, email

Description automatically generated

The weakness of this dataset, as follow:

1. The country name is not a country name, sometimes its region such as NATO or Euro. This makes problem for cross country data analysis.
2. There are too many indicators, and the indicator values is in row. This make it hard to create cross country and timeseries analysis based on the dataset. For example, I cannot create a scatter plot for cause variable and response variable based in this dataset. I also cannot create a heatmap based on the dataset.
3. The dataset is too big. Even in phyton, processing it takes time.
4. Therefore, for the dashboard, I need to create a selection and transpose the dataset, as follow:

Table

Description automatically generated

1. Based on my experience, this data is often used in economic research. Therefore, in order to create a future help for researcher and policy makers, I create a function that helps to select the targeted countries and specific time period using countryselection() function.
2. The function needs the specific target year for snapshot or period for timeseries.
3. User needs to write the name of selected country into a file named (listofcountry.txt) in the same folder. The function will automatically read the file and subset the dataset based on the countries in the file.
4. I also create a function named transposetransformation(). This function will pick the specific indicator. The function needs a specific period as input and txt file that contains the specific indicators. This function will also transpose the table.
5. Indicator initially as row become column
6. Year initially as column become row
7. Those functions live in FinalProject.py. User needs to use the function in correct order to make it works and create a beautiful time series data from multidimensional world bank data. This also will make the dataset smaller and easier to process. The final result of the table for the dashboard as follow:

Table

Description automatically generated

With this data structure, I can analyze a snapshot for cross country analysis or timeseries analysis, and the function makes the selection of countries, indicators and years easier. Therefore, Users can work with smaller dataset for research or analysis.

1. Data for the heatmap is combining two datasets, one from the choropleth map dataset, as follow:

A picture containing text, scoreboard, monitor, cabinet

Description automatically generated

Then, I join it with my datasets, as follow:

