**CA682 Data Management and Visualization**

|  |  |
| --- | --- |
| **Name** | HARSHALI PATIL |
| **Student Number** | 18212797 |
| **Program** | MCM |
| **Module Code** | CA682 |
| **Assignment Title** | Data Visualizations |
| **Submission date** | 16/12/2018 |
| **Module Coordinator** | Suzanne Little |

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations set out in the module documentation. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study. I have read and understood the referencing guidelines found recommended in the assignment guidelines.

Name: HARSHALI PATIL

Date: 16/12/2018

**VISUALIZING RELATION BETWEEN LIFE EXPECTANCY AND GROSS DOMESTIC PRODUCT**

# **INTRODUCTION**

A country’s economy depends on various factor, which can be shown using different development indicators. Here, Gross Domestic Product is used to show the development of various countries.

Gross Domestic Product (GDP) is the measure of market values, final prices of the goods and services produced in the geographic boundary of any country in a specified time period. It reflects the nation’s total economic value. It considers the final value of a product, but not the intermediate good that go into it. To cite an example, GDP only counts the final value of sugar, but that does not mean that the contribution of the factory is considered as a final price. Final sugar price would be the sum of all the contributions made to produce sugar from a raw sugarcane. Well, GDP is expressed as the comparison to the previous quarter or a year. For example, if a nation’s GDP value for 2016 is 5% increased, it means that the economy of that country is grown to 5%.

**TYPES OF GDP**

## **GDP CURRENT:**

It is the gross market value of the final goods produced in a country during a year. It is commonly used to determine the economic growth of a country. Currently, it is measured in US dollars; currencies of all the countries are converted into US Dollars. In GDP current, market exchange rates are considered for conversions. The highest GDP current value in the world is for United States.

## **GDP PURCHASE POWER PARITY (PPP):**

It is useful to compare the differences in living standards among countries. Like, a pizza costs more in London than in Rwanda. It tells how expensive it is to live in a nation. It compares currency of different countries using market basket of goods approach. PPP is measured by comparing the prices of identical items of different countries. It provides more accurate rates than the market exchange rates. It is difficult to calculate as it does not consider the quality of goods and services among countries. PPP rates are constant over the long run.

## **GDP PER CAPITA:**

GDP per Capita is the best approach to calculate Gross Domestic Product between nations. It is calculated as the ratio of Gross Domestic Product to the population of that country. It reflects the country’s living standard. It is an important indicator of economic health of a country. GDP per capita is the average amount of money spent per person. It does not consider the income of the individuals residing within the country.

Life Expectancy is increasing over the decades. It was noted that the pre-modern or the poor world had the life expectancy of only 30 years. It can be seen that the early industrialized nations have higher life expectancy than the rest of the world. Richer Nations have higher life expectancy and the poor nations accounting lower life expectancy; probably due to lack of facilities provided to the individuals of that country. This project visualizes the GDP Current, GDP PPP and GDP per capita of the 217 countries. Moreover, this project aims to show the correlation between the life expectancy and the GDP of a country.

# **DATASET**

In this project, various publicly available authenticate datasets has been used. As the available datasets contained noisy information, the datasets were cleaned using Python to get the desired refined dataset.

**SOURCES:**

1. **WorldBank Dataset:**

The World Bank recognizes that transparency and accountability are essential to the development process and central to achieving the Bank’s mission to alleviate poverty. The World Bank’s commitment to openness is driven by a desire to foster public ownership, partnership and participation in development from a wide range of stakeholders. The World Bank shares its knowledge freely and openly. Statistics and data are a key part of that knowledge and are easily accessible on the web for all users. The World Bank provides free and open access to a wide range of datasets about development in countries around the globe, together with other datasets cited in the data catalog.

**Source Link :** <http://databank.worldbank.org/data/source/world-development-indicators>

Following are the datasets used for visualization which are accessed from WorldBank Dataset:

* **GDP (Current US$):** This data was extracted from WorldBank Dataset, which consists of GDP current values of 217 countries for year 2016.
* **GDP(PPP):** This data was taken from WorldBank Dataset, which consists of GDP per capita values of 217 countries for year 2016.
* **GDP per capita (current US$):** This data was taken from WorldBank Dataset, which consists of GDP per capita values of 217 countries from year 1960 to 2016.
* **Life Expectancy:** This data was taken from WorldBank Dataset, which consists of life expectancy in 217 countries from year 1960 to 2016.

# **PROCESS**

The above-mentioned datasets were downloaded in CSV format and then cleaning, and merging was performed over the datasets in order to get the refined dataset.

Python was used to clean and slice dataset using Pandas and Numpy libraries. The description of the both the steps are given below:

Cleaning: Python was used to clean the datasets. Firstly, Datasets were read in Python using Pandas. Secondly, there were some unwanted rows and columns, so, deleted those rows and columns using drop function. Also, the column names were not in proper format so renamed the column name using rename function. Lastly, replaced the “..”, which shows the missing values in the dataset, with 0.

Slicing: For the Line graph, dataset contained the Life Expectancy and the GDP per Capita of 217 countries from 1960 to 2016. Using Python, this dataset was sliced into two different data frames.

## **TOOLS USED FOR VISUALIZATION**

1. **PLOTLY: Plotly is a data analytics and visualization tools to create, modify and publish interactive graphs on the web. It provides Python libraries to make interactive graphs. Plotly offline provides the graphs locally.**
2. **IPYWIDGETS:** Ipywidgets are HTML widgets to provide interactive controls on Jupyter Nootebook and IPython Kernel. It enables the user to gain control of the data to visualize the change in data using slider, dropdown etc.

# **PLOTS**

1. **CHOROPlETH MAP:** The first visualization is shown using Choropleth map for world. It represents the Gross Domestic Product of each country around the globe. Here, countries are colored, and the higher values are represented by darker/stronger color while the density of color reduces as the GDP values decreases. Colors were chosen from color pallet generator called “Coolers” (<http://coolers.co>) and meet standard guidelines set by WCAG2(Web Content Accessibility Guidelines.

**To make the graph interactive:**

**Dropdown menu from Plotly is used.** Dropdown menu allows to visualize the data change according to the choice presents in the dropdown menu. Specifically, GDP Current, GDP purchase Power Parity, and GDP per capita of 217 countries for year 2016.

A close up of a map

Description generated with high confidence

Figure 1: Gross Domestic Product, 2016.

1. **LINE GRAPH:** Line graphs are well-suited for comparing different datasets. In this graph, Life expectancy is compared with the GDP per capita of the 217 countries from year 1960 to 2016. This graph is a dual y-axis graph; one for life expectancy and another for GDP per capita in US$. On x-axis, years are taken from 1960 to 2016. Markers are used to have the clear view of data points. Colors were chosen from color pallet generator called “Coolers” (<http://coolers.co>) and meet standard guidelines set by WCAG2(Web Content Accessibility Guidelines.

**To make the graph interactive:**

**Widgets from ipywidgets are used. Particularly, Dropdown widget is used to select a country out of 217 countries.**

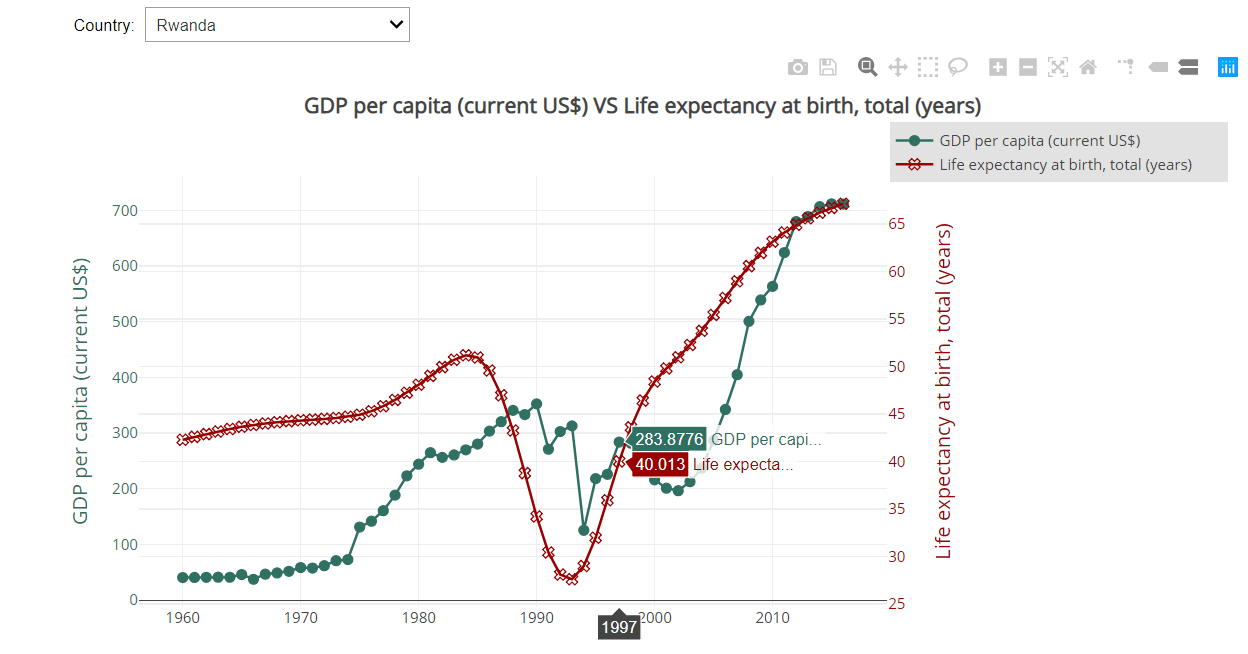


Figure 2: GDP per Capita (current US$) VS Life Expectancy

# **RESULTS**

It can be seen from the graph that United States has the highest GDP Current in 2016 around the world i.e. 18.62448 Trillion. China, Japan, Germany, United Kingdoms, France, India, Italy, Brazil, Canada, South Korea is ranked in descending order after US.

GDP purchase power parity has data values in trillion and the china overtook United States to become highest in GDP PPP with around 21 Trillion US$. Top 10 countries having highest GDP PPP value in descending order: China, USA, India, Japan, Germany, Russia, Indonesia, Brazil, United Kingdom, France.

GDP per Capita shows the economic performance of the country and the country’s standard way of living. It can be noted from the graph that Luxembourg has the highest GDP per Capita. Singapore, Ireland, Norway are in second, third and fourth position respectively.

From the line graph, it is observed that the Life Expectancy varies with the GDP per capita. Richer Countries have higher life expectancy as they can provide good health facilities while the poor countries have lower life expectancy. Countries like Rwanda, Niger, Uganda has lower GDP per capita, thus are the poor countries having 65 years of life expectancy. Countries having higher GDP per capita like Luxembourg, Singapore, Ireland has the higher life expectancy of around 82 years.