School of Science, Engineering and Environment.

M.Sc. in Data Science

Assessment Title:

Socioeconomic Analysis of selected Developed and developing countries

Name: Stella Awoyomi

Student ID: @00655282

Table of Content

1. Introduction

In the last ten years, there has been a mixed development in the economic situation of some countries of the world. This trend has benefited some countries, and some are declining. A country's economic situation will impact its social activities, which must be studied carefully. In order to achieve the highest level of human development possible, it serves to sustain the nation's and its citizens' social and economic well-being. Also, to understand the gap that developing countries need to fill to improve the livelihood of their citizens. This report aims to study and compare the socioeconomic development of selected developed countries against some developing countries and observe the fast-developing and struggling ones.

The objective of the report is to compare the gross domestic product of the selected countries from 2011 to 2022. Also, examine which country within the chosen time that improvement in percentage increase in gross domestic product. Comparing the countries exporting power and its contribution to gross domestic product. World population growth has always been on the rise; the report will examine its effect on inflation and gross domestic product. Also, study the relationship between inflation and life expectancy. This report will use a graphical presentation using a dashboard showing the relationship briefly. The dashboard will display the output and the finding from the dataset.

This report is structured in sections; section one is the introduction, and section two will give background on the research, focusing on the ethic of building a dashboard. Sections three and four explained the exploration of the dataset, and investigation of data workflows,

respectively. The last two sections will focus on findings, discussion and conclusion.

2. Background Research

Monitoring, regulating, and improving an organisation's operations are based on visualising organisational performance. Dashboards are an effective tool for combining pertinent information in a single view that provides a graphical summary of the current condition (Staron, 2012). A "dashboard" is an intuitive real-time user interface that aids decision-making by displaying graphical representations of the current status and historical trends of an organisation's Key Performance Indicators.

A user-centred, goal-centric design that adheres to best practices for dashboard design and appropriate data visualisation makes dashboards an effective tool for communicating data and other information. Even though each dashboard is unique and has its objectives, constraints, and requirements, adhering to these fundamental principles can help designers produce stunning designs irrespective of the circumstances. Currently, end-users and analysts are just two of the actors involved in the life cycle of a dashboard (Chiang,2009). End users participate and offer feedback to the business analysts who design personalised dashboards to fulfil user demands in user-centred BI (Foody, 2009). This input is provided at various phases of dashboard setup and design, requiring extensive discussion between analysts and end users to specify functional requirements and ensure a satisfying user experience.

There are many proposed methodologies for building a dashboard. One is the methodology of Hariyanti, which gave three core aspects a dashboard must possess. The aspect is data, personalisation and collaboration (Hariyanti, 2008). Also, there is Tufte's data – Ink Ration; Tufle defined data-ink ration as the percentage of a graphic's ink used for the sequence display of data. Tulle proposed five principal procedures in graphic design (Tufte, 1983).

Furthermore, an excellent place to start when creating an advanced graphical user interface was suggested by Shneiderman, the "Visual Information Seeking Mantra", which puts Overview at the top, after which Zoom and filter come after, then the details on demand (Shneiderman, 1996). Another model is Munzner's nested model for visualisation design; this method suggests a four-level model for visualisation design and evaluation. Characterising the issues and data in a particular domain at the top level, translating those into abstract operations and data types at the second level,

designing visual encoding and interactions to support those operations at the third level, and finally, designing an algorithm to execute that design automatically and effectively at the fourth and lowest level (Munzner, T. 2014). The last framework which combines all the mentioned methodologies above is the 5S dashboard design principle framework. This framework is built on five Japanese words translated as "sort, set in order, shine, standardise and sustain (Gapp et al., 2008)".

The five dashboard design principles that will be employed in this contest brief are below:

1. Focus on end users

Most dashboard design principles centre on the requirements of the final users. Decide what data the user wants to study before you begin the design process—select key performance indicators in collaboration with users, clients, or stakeholders. The data you must present in the dashboard influences the remaining design steps. After establishing the KPIs, you may choose the appropriate dashboard and visualisation tools.

2. Using the right dashboard type

The proper type of dashboard should be chosen as part of the dashboard design guidelines followed by seasoned designers. You might need a KPI dashboard if the dashboard is meant for senior executives and decision-makers. An operational dashboard may be necessary for managers who watch what employees do.

3. Give Quick Access to Useful Information

The best dashboard design practices include providing users with quick access to the most crucial information. The "five-second rule" is what is widely employed. After the dashboard loads, all pertinent data should be accessible within five seconds.

4. Choosing the appropriate data visualisation type

One of the most frequently ignored dashboard design elements is the necessity to use the appropriate visualisation tools. Designers frequently overdo it when choosing different graphs, tables, charts, and maps. Refrain from filling the screen with several charts and tables when using visualisation techniques. To ensure that the dashboard has a uniform appearance, use a limited number of visualisation kinds.

5. The design should be simple and easy to understand with storytelling

One of the most crucial dashboard design tenets is to make the interface as straightforward as possible. Make it as simple as possible for users to examine the data displayed on the screen.

From the same data collection, various users can draw different conclusions. Finding a balance between author-driven components (which provide narrative structure and information) and reader-driven aspects is necessary.

3. Exploration of dataset

The dataset used for this report is secondary data collected from World Development Indicator (World bank data bank). Eight variables were used to compare the socioeconomic situation of some selected countries. The variables are:

- Gross Domestic Product per capita (USD) (GDP per capita)
- Gross Domestic Product growth percentage per annual
- Export of goods and services in USD
- Export of goods and services percentage contribution to Gross Domestic Product.
- The inflation rate, consumer price in USD
- Life expectancy at birth (years)
- Population growth in percent.

Also, ten years was chosen to compare the performance of the countries in a decade. Developed countries and some developing countries across continents were chosen. Twelve countries were considered: the United States of America, China, Japan, Brazil, Germany, France, India, Australia, the United Kingdom, South Africa, Nigeria, and Egypt.

The data contained some missing values, but because the data we are using is time-dependent and cannot be filled with mean values, the forward fill method was employed to cater for the missing values. Also, the data downloaded was in the form of long data converted to a comprehensive dataset for visualisation. All the data used were integers except for country and year, which were changed to a text data format. The dataset rows represent entities with a unique year and country combination. The columns break down the variables or indicators used to compare the countries' socioeconomic status. Finally, due to the nature of the dataset collected, there were two tables in the modelling process with many to many relationships. The economic table and the social index table. In order to make a good connection between the two tables, a new column was created for the primary key in each table by

combining the year and the country code column in the power query editor.

4. Investigation data Workflows & Proposal for Design of Dashboard

The workflow of this report consists of five major steps that will make a dashboard design have a strong connection with the objective of the report. Also, the workflow helps in building a dashboard that is the representation of the objectives. How the workflow assisted in getting the solution to the problems will be elaborated. The workflow is represented infigure 1 below.

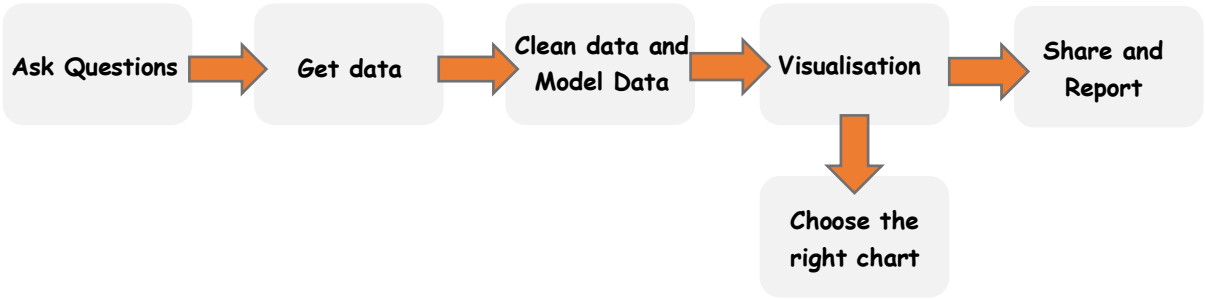

Figure 1: Data workflow

Ask Questions

In any data visualisation report, there must always be a question and problem that induced the process. This report answers the question of which countries have the best economic among the countries selected. Which country has the highest percentage of gross domestic product per capita growth? Also, what is the relationship between inflation on consumer goods and population growth? Which country has the highest exportation of goods and services and contribution to the country's gross domestic product?

Get Data

This step answer HOW the question will be answered. The data collected are the socioeconomic indicators dataset that will assist in achieving the objective of the report. The data were downloaded in an excel format from the World Development Indicator (World bank data bank). The economic and social development indicators are from two comma-delimited files (CSV). The dataset collected is the indicators used to measure a country's socioeconomic development. The datasets were imported into power query from

two excel CSV files. The data span over ten years, from 2012 to 2021.

Data Cleaning and Modelling

Data cleaning is an important aspect of modelling. Missing values and wrong data labels might cause an error during visualisation. The datasets imported have missing values which required good tidiness and cleaning before visualisation. Power Query is a powerful tool used to clean and transform the dataset. The dataset has been time-dependent; we visually checked for the missing data and used a forward-fill filling method to cater for the missing data. Also, the data type needed to be corrected; some currency data were informed of the whole number, which is a wrong assignment. Firstly, the data that should be represented by percentage (Gross Domestic Product growth percentage per annual, Export of goods and services percentage contribution to Gross Domestic Product, and Population growth in percent) were changed from the whole number to percent. Also, the dataset in United state dollars (USD), like Gross domestic product per capita, Inflation rate, consumer price in USD, and Export of goods and services, was changed from the whole number to USD. The data initially does not have a specific key to join two tables together for modelling. Two columns, which are year and country code, were combined to create a new key to connect the two tables. The table model is shown below:

Visualisation

Even though you have already started by creating your dashboard, it is crucial to understand which chart style will work best with which kind of data. When used improperly, data visualisation can be deceptive. Visualisation can only be used in time-dependent datasets. The report compares 12 countries' socioeconomic development using seven indicators: Gross domestic product and total importation USD and its contribution to GDP, Gross domestic product growth, inflation, population growth and life expectancy. The Gross domestic product of twelve countries in ten years will best be visualised using a small multiple-line plot which shows a trend over time (Steve et al., 2017). Furthermore, bubble plots or scatter plots are the recommended plot used in checking the relationship between three socio-economy indicators like Gross domestic product growth, population growth and life expectancy in a specific year. The effect of inflation on countries' life expectancy was visualised using a line and stacked column chart; this chart shows the relationship through a line graph in combination with a stacked column by comparing the counties side by side and taking the average of the ten years (Schwabish, 2021).

The importation strength of each country and its percentage contribution to the country's Gross domestic product per capita was compared using a line and stacked column chart. The chart has county name on the x-axis, total importation in USD on the y-axis, and the trend represents the exportation contribution to Gross domestic product per capita. The chart was used in the dashboard design. The dashboard consists of a title section and two slices based on year and country. Also, the economic indicator comparison and the social activities chart. To make this dashboard interactive, the slicer base on year and country was used and placed at the top right corner of the dashboard. The slicer is in the form of a dropdown which will help the user to view the annual representation of each indicator.

Share and Report

The final stage of a dashboard is sharing and reporting the solution to the stakeholder. The dashboard is only relevant if it answers the questions before data collection. Sometimes it may require deployment to the website or mobile application. This dashboard designed for tis report will be published on a website.

Interpreting the Dashboard Charts

The GDP per capita by country name and years using a multiple-line plot chart shows that Australia had the highest GDP per capita at the beginning of the century, which the United States overtook. In 2021, the United State boosted the highest GDP per capita after the Covid-19 pandemic, followed by Australia and Germany. The countries with the lowest GDP per capita are Egypt, Nigeria and India. This shows that developed countries have better GDP per capita than developing countries. The GDP per capita growth average chart shows that China is the fasted growing economy, followed by India and Egypt in second and third, respectively. Nigeria, Brazil, and South Africa have declined the GDP per capita growth average over the past ten years.

Furthermore, China has the highest Export of goods and services, contributing 20 percent to their GDP per capita. Germany's Export of goods and services contributes up to 50 percent to their GDP per capita. Also, it was observed that there is a positive relationship between population growth and inflation in consumer prices in 2021. Moreso, Nigeria recorded the highest inflation rate between 2012-2021, followed by Egypt and India, with Japan, France and Germany having the lowest inflation rate, respectively. Life expectancy and inflation show a negative relationship. Countries with lower inflation rates have a higher life expectancy. This means that inflation causes a citizen to work overtime to meet the cost of living during inflation, reducing life expectancy.

5. Discussion

A goal-centric and user-centred design that adheres to best practices for dashboard design and appropriate data visualisation makes dashboards an effective tool for communicating data and other information. There are many proposed methodologies for creating a dashboard. These methodologies have something in common, which is a user and goal-centred interface. The methodology employed for this report is the 5S dashboard design principle framework. The framework is based on the 5S, from Japanese words translated as "sort, set in order, shine, standardise and sustain (Gapp, 2018)". The five dashboard principles designed that will be employed for this report mainly focus on end users. It is

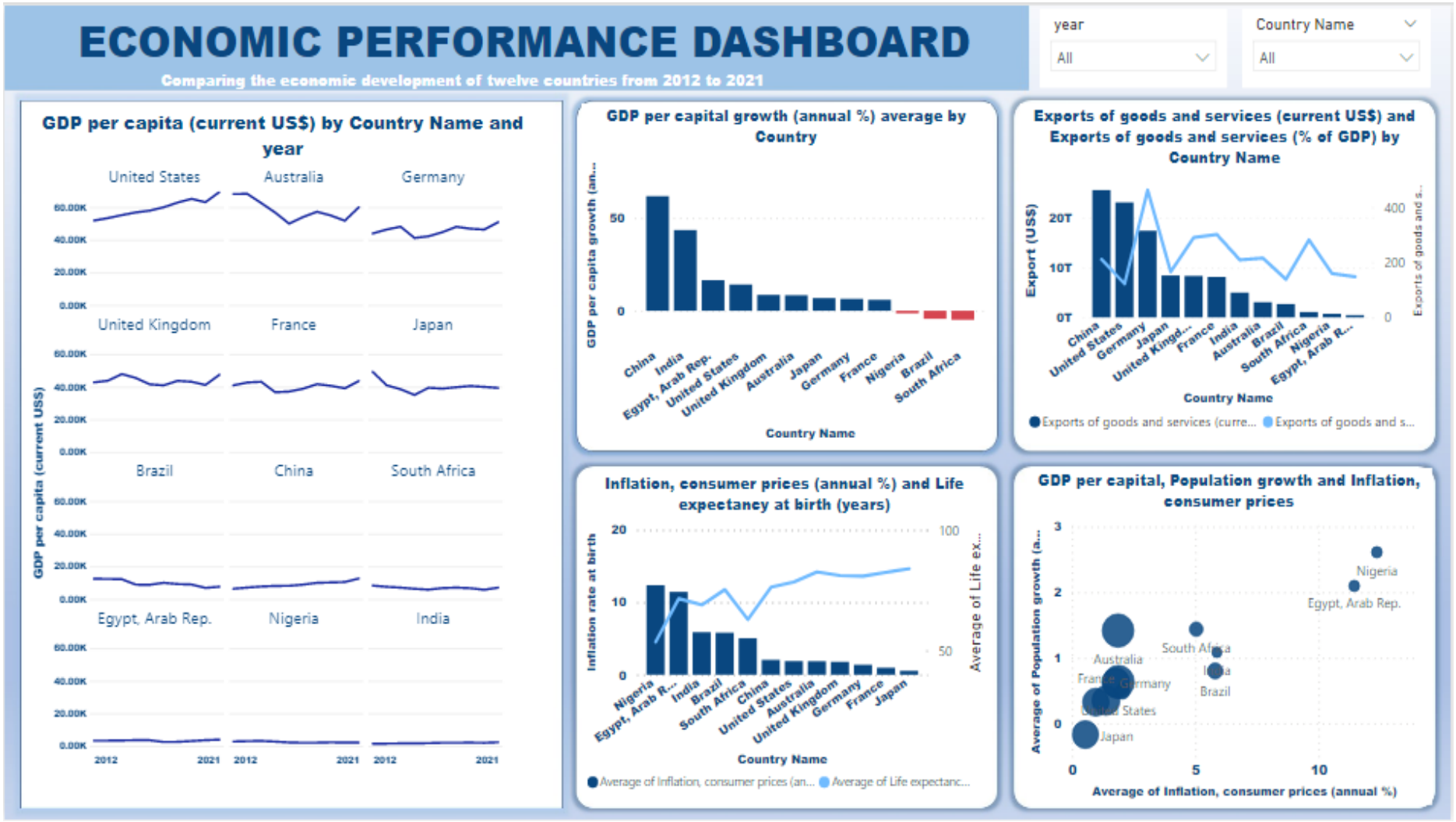

Figure 2: Dashboard

incumbent to decide which data users want to study. Also, using the proper dashboard type is another part of the dashboard design.

Furthermore, the dashboard must provide quick access to the most crucial information within five seconds of loading the dashboard. Moreso, Using the appropriate visualisation tools in designing the dashboard element. The dashboard should be simple and easy to understand with storytelling.

Workflow is a crucial aspect of dashboard design. A dashboard foundation must answer a specific question on a particular problem that needs to be solved. The questions will lead to collecting the needed data to assist in solving the problems. The data used here are the socioeconomic development indicators for ten countries. The countries consist of developed countries and developing countries. Data collected are only sometimes clean; it is important to browse through the dataset to observe its irregularities. Sometimes, the missing value might be present in the data. An invalid data type is another issue that must be examined in data cleaning. Sometimes the data needs to be reshaped for the dashboard design to be possible. The dataset used for the dashboard in this report was not free from missing data and wrong data types. The missing values in the data were filled using forward fill because of the nature of the data time series. Also, the right data type was assigned to each variable.

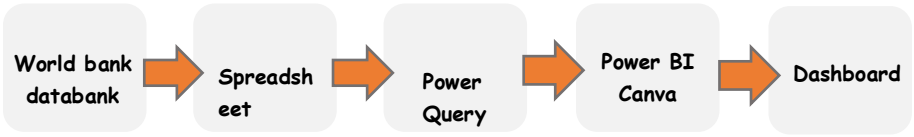

Figure 3: From Data to Dashboard

After the data cleaning, the tables were joined using the common key in the two tables in modelling. Different datasets with different choices of visualisation charts to choose from. It is well known that bar charts are good for comparing categorical variables and continuous variables. Also, line plots are good for data sets that are time dependent. A map graph is also suitable for research and data that are location-depending. Sharing the designed dashboard with the user in a way

that will be self-explanatory. They are also deploying it to the website and maintaining the data flow. A shared dashboard must be solution-driven, or else the dashboard will be misleading by portraying wrong information and be irrelevant to the problem. The final dashboard workflow during design is given in figure 2 above. The final dashboard consisted of seven parts; The first part is the title section, followed by the year and the country slicer. Also the

third part is the multiple-line graph showing each country's GDP. The next graph, which is a bar plot, shows the average GDP growth in the past ten years. Also, the fifth section shows countries' Exports of goods and services compared to their contribution to the countries' GDP per capita. The sixth graph shows the countries' inflation on consumer prices. It was compared to their life expectancy, which shows that Nigerians recorded the highest inflation rate in 10 years with the lowest life expectancy. The last line and stacked column chart show the relationship between population growth and the inflation rate of the countries in relation to GDP per capita.

6. Conclusion

In conclusion, this report aimed to create a dashboard that will compare the socioeconomic situation of developing and developed countries. The dashboard design shows the country with the highest GDP per capita from the past ten years. Also, the country has had the highest average GPD growth in the past ten years. Also, in the dashboard, we compare the relationship between population growth and inflation in relation to the country's GDP per capita. The countries' inflation on consumer prices was compared to their life expectancy. Countries' Export of goods and services as compared to their contribution to the country's GDP per capita.

This final dashboard was designed based on the methodology specified. There are some methodologies proposed by researchers and some principles which should be followed. The methodology employed in this report is the 5S dashboard design framework. Also, the principle used focused on five points which are focused on end users, using the right dashboard type, giving quick access to useful information, choosing the appropriate data visualisation type, and making the design simple and easy to understand. The workflow is crucial in the dashboard design process. Asking questions, getting the data needed to answer the question, data cleaning and modelling, visualisation and storytelling. These were the foundations of the designed socioeconomic dashboard.