Introductory Guide on Power BI with Demonstration

Abstract

Power BI is a cloud-based business analysis and intelligence service by Microsoft. It provides enterprises with strong data analytics and visualizations to have a deeper understanding of data from various sources. Although Power BI is a powerful tool, users should know in which use cases Power BI is the most appropriate weapon for them over other tools and how to mitigate the limitation of Power BI in dealing with large datasets. A tutorial of how to set up Power BI Desktop and an illustrative example are essential for users to have a sense of how to apply Power BI in the real world to visualize, analyze and extract insights from data. A specific demonstration using the World Bank Indicators dataset revealed the strengths and the weaknesses of Europe from the period from 2000 to 2010. Europe achieved the highest GDP per Capita, spending the highest budget on the health care system, as a result, gaining the highest life expectancy at birth and the lowest mortality under 5 years old. However, it was exposed to some weaknesses including an economic downturn trend, aging population, and low birth rate. Based on the strengths and weaknesses, users could give some recommendations to the authority, such as they should identify reasons for economic downturns to apply appropriate solutions; as well as encourage maternity to resolve aging population issue, enlarging the labor force and boosting the growth of the economy.

Introduction

We live in the world of data deluge, and data analytics become a unique selling proposition (USP) of a business if it wants to survive and prosper over its competitors. The purpose of data analytics is to gain insights from raw data and make educated decisions. Data is much more valuable when it is visualized by graphs and charts. Visualizations can uncover trends, patterns, the underlying structure of data or connections in data, and even surprising findings that data scientists may never think of without data visualization. Moreover, visualization assists readers in grasping ideas quickly since an image is worth a thousand words. Visualization is one of the most effective tools for all kinds of storytelling, especially for complicated stories in science. However, visualization

has not been largely recognized for its crucial role in analyzing data and communicating results to readers. It is still considered as icing on the cake and the last thing that scientists or researchers do. Therefore, apart from the data analyzing data skills, data analysts should empower themselves with good visualization skills to be a more effective data analyst. The goals of this project are to communicate the importance of visualization, as well as introduce one of the most powerful tools for data analyzing and visualization: Power BI. Microsoft offers three types of Power BI platforms: Power BI Desktop, Power BI Service, Power BI Mobile, but this project would brief mainly on Power BI Desktop. Although Power BI has three main functionalities, including Data Shaping, Data Modeling, and Data Visualization, this project would only concentrate on Data Visualization other than Data transformation or data modeling.

Methodology

Firstly, basic knowledge about Power BI will be introduced, including a brief introduction of its main functionalities: Data Shaping, Data Modeling, and Data Visualization. Secondly, the cases in which users should choose Power BI over its biggest competitor, Tableau are also recommended. Thirdly, some techniques to deal with large datasets/big data in Power BI are introduced as well. Lastly, there is a tutorial on how to setup Power BI Desktop and for specific demonstration, Power BI would be applied in a specific dataset to show how to use common features such as reading in data, creating relations between datasheets, perform some basic data shaping (where it is required) before visualizing data, and how to create a report or an interactive dashboard, and gain some insights/findings with recommendations. The dataset used for the demonstration is the World Bank Indicators dataset. This dataset consists of data about Transit (Car or Railways), Population (by different age groups and Mortality rate), Health (Life Expectancy, Health Expenditure), Business (Mobile phone subscribers, Internet users), and Finance (GDP and GDP per capita) of all countries in the world from the year 2000 to the year 2010.

Introduction of POWER BI and its main functionalities

Power Bi is well-known as a collection of business intelligence and data visualization tools developed by Microsoft. Datasets from multi-sources imported in Power Bi would be visualized and analyzed by generating intuitive reports, dashboards, and apps to extract insights and share them across your organization for efficient decision-making. Power BI impresses users with ease of use thanks to its drag-and-drop features and self-service capabilities.

Power BI has three main functionalities: Visualization, Data shaping, and Data Modeling:

1. <u>Visualization</u>

- Users can create reports using a wide range of visualizations, such as Stacked bar/column chart, Clustered bar/column chart, 100% stacked column/bar chart, Line chart, Area chart, Stacked area chart, Ribbon chart, Waterfall chart, Scatter chart, Pie chart, Donut chart, Tree map chart, Map, Filled map, Funnel chart, Gauge chart, etc. Power Bi also offers a library available for custom visualizations.
- Power BI offers more than 3500 data points for drilling down the dataset and visualize data.
- Users then can create interactive dashboards from a collection of visualizations as tiles to communicate meaningful information or insights into data. The dashboards are single pages from the reports, and they are shareable as well as printable.

2. Data Shaping

Users can shape data in Query Editor by providing step-by-step instructions to adjust the data as it loads (the original data source isn't affected). These steps are recorded sequentially in the Query Setting pane, under Applied Steps.

- Shaping data or transforming the data includes renaming columns or tables, transform a data type (e.g. changing text to numbers), removing rows/columns, setting the first row as headers, etc.
- Combining data includes connecting two or more data sources, shaping them as needed, then consolidating them into a useful query.
- Combining queries includes merging (adding one or more columns to an existing query) and appending (add one or more rows of data to an existing query).

3. Data Modelling

Users can use Power Bi Data Modelling to connect multiple information sources employing a relationship and produce visualizations on those sources.

- With the modeling feature, users can build custom calculations on the present tables or outline new metrics to look at specific segments of data, perform custom calculations for those metrics, and visualize these new measures for straightforward modeling.

- Users can produce calculated columns by combining 2 or more associated existing columns to outline a replacement metric or mix 2 columns to form one new column. Users can even produce a calculated column to determine a relationship between the tables and use that column to form a relationship between 2 tables.

- Users can conjointly produce a replacement calculated table in information modeling employing DAX expressions.

Use cases of Power Bi vs Tableau:

Power BI	Tableau
For both non-experts and experts in analytics: Power BI is easy to use as it bases on drag- and-drop and intuitive features.	For experts as data analysts as the interface is not quite intuitive. People may need data analysis to clean and transform data into visualizations.
Appropriate for small businesses with limited financial and human resources, especially if they already invest in Microsoft products.	Appropriate for medium and large companies that have supporting IT resources and prioritize speed and capabilities of data analytics.
For both companies with data warehouses with structured data and companies with less structured databases that can be reprocessed in DAX.	For companies that already have a team of data scientists, structured data warehouses, an elaborate data preparation process
Users' main goal is data manipulation and reporting.	Users' main goal is data visualization and scaling
Users' companies do not have a big data warehouse with a well-defined ETL process since Power BI can handle a limited volume of data.	Dealing with big data since Tableau works best when there is vast data in the cloud
Users are familiar with or prefer using M and MDX languages for measures and dimensions	Users are familiar with or prefer using DAX programming language for calculating and measuring columns.
Users want to share users' work with others who don't have Power BI accounts by using Power BI Embedded to create embedding reports.	An embedding report is a big challenge
Users need connectivity to a wide variety of complex data sources	Users just want to base on visual aesthetics alone
Users want a quick analysis with out-of-the-box dashboards or connectivity to popular technologies such as Google Analytics, email platforms, and various Microsoft products, etc.	Users want out-of-the-box integrations and connectivity to large tools and connections.

Users have to do the same actions every time, Power Bi allows users to make templates for processing	Creating a template does not work in Tableau as it is designed for already preprocessed data.
Users who don't know how to create a graph can use Natural Language Q&A tool in Power BI to write questions, and they will be provided with the content and answer quickly.	If users want to create a story graphically
Users only work with data models in the software	Users only work with flat tables

How to deal with large datasets/ Big Data

There is a 1 GB limit per dataset that is imported into Power BI Pro and 10GB for Premium. Power BI Desktop uses Vertipaq storage engine to compress and optimize import models 10 times then store them to disk, which can achieve 20% additional reduction in size. However, for large models or models growing large over time, users should apply the following techniques for data reduction:

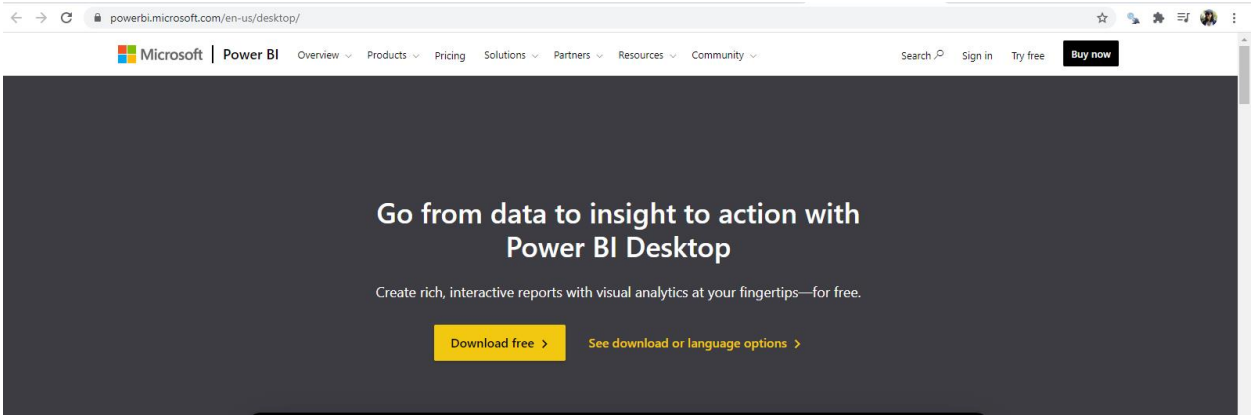
- Vertical filtering: Remove unnecessary columns which do not serve the purpose of reporting and model structuring. Users should design models with exactly the required number of columns. If the requirements change over time. It is easier to add than to remove columns later since removing columns can break reports or model structures.
- Horizontal filtering: Remove unnecessary rows
  - Users can filter by entity which involves importin a subset of source data into the model.
  - Filtering by time which limits the date rows imported into the model.
- Group by and summarize data to reduce a model size before importing
- Convert source text data into numeric values where applicable. This will result in a significant data reduction since text and other non-numeric data use hash encoding, which requires the storage engine to assign a numeric identifier to each unique text value in the column and store them in the data structure. Users need to do a hash lookup during storage and querying. Meanwhile, numeric column data use value encoding, which can achieve the highest storage optimization.
- Create custom columns in Power Query as computed columns (defined in M) instead of calculated columns (defined in DAX) to achieve greater load efficiency. If the source is a database, the calculation can be defined in SQL statement, using the native query language, or can be materialized as a column in the data source.

- Avoid loading Power Query queries to the model when importing data by disabling Power Query query load.
- Disable the auto date/time option to avoid creating a hidden auto date/time table for date columns since these tables will increase the size of data.
- Set the Storage Mode property of large tables to Import or DirectQuery to reduce model size since the 1-GB dataset limitation doesn't apply to DirectQuery.

For storage of large datasets, users must purchase Power BI premium to enable Large dataset storage format setting to able to use datasets that are larger than the default limit. Data will be stored in a highly in-memory cached for optimized query performance, enabling fast user interactivity.

How to set up Power BI Desktop

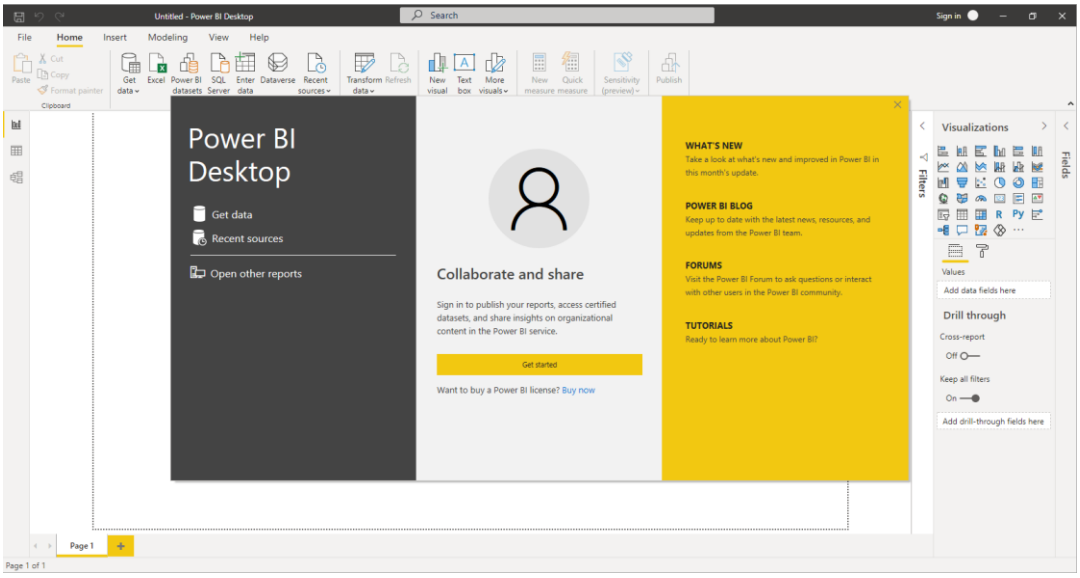
Step 1: Go to the link: https://powerbi.microsoft.com/en-us/desktop/
And select '**Download free'**



Step 2: On the Microsoft Store page, select **Get**, and follow the prompts to install Power BI Desktop on users' computers.

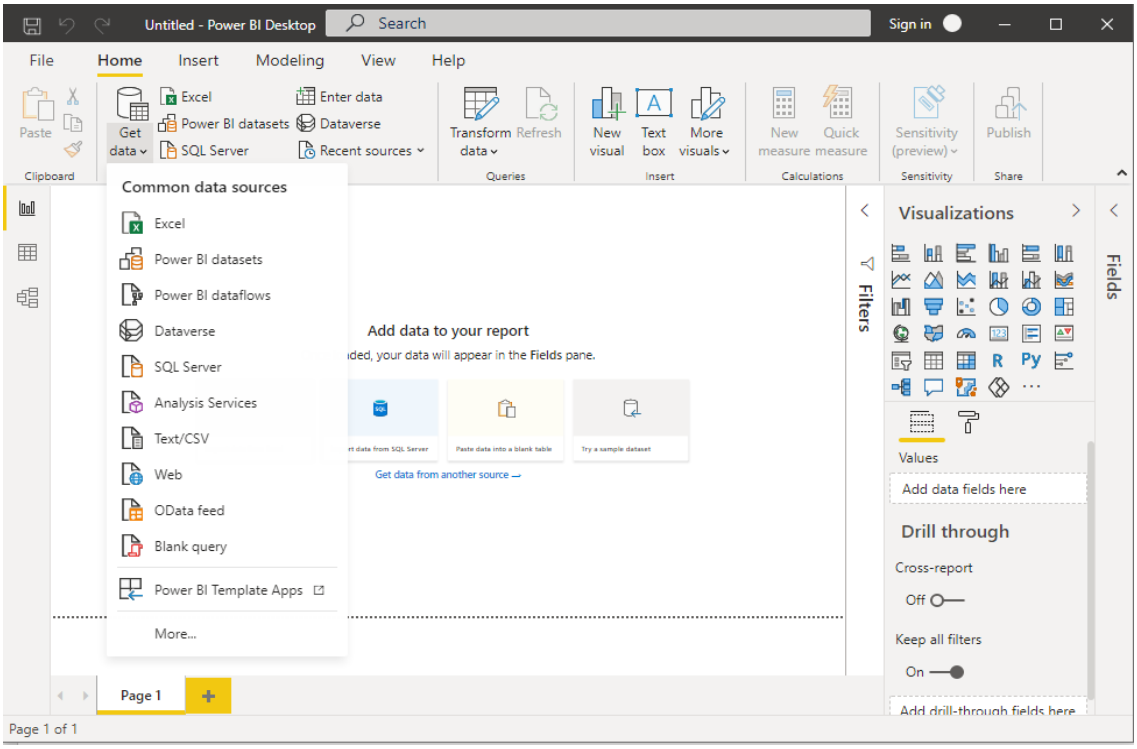
Step 3: Start Power BI Desktop from the Windows Start menu or the icon in the Windows taskbar.

The first time Power BI Desktop starts, it displays the Welcome screen. Just close it.

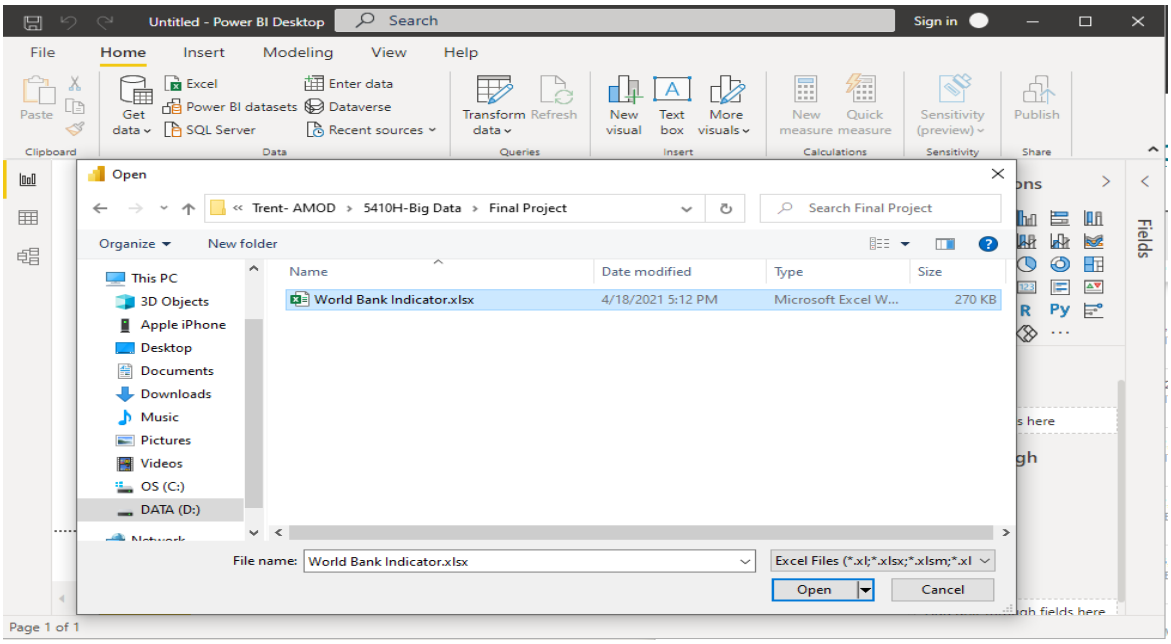


Specific demonstration:

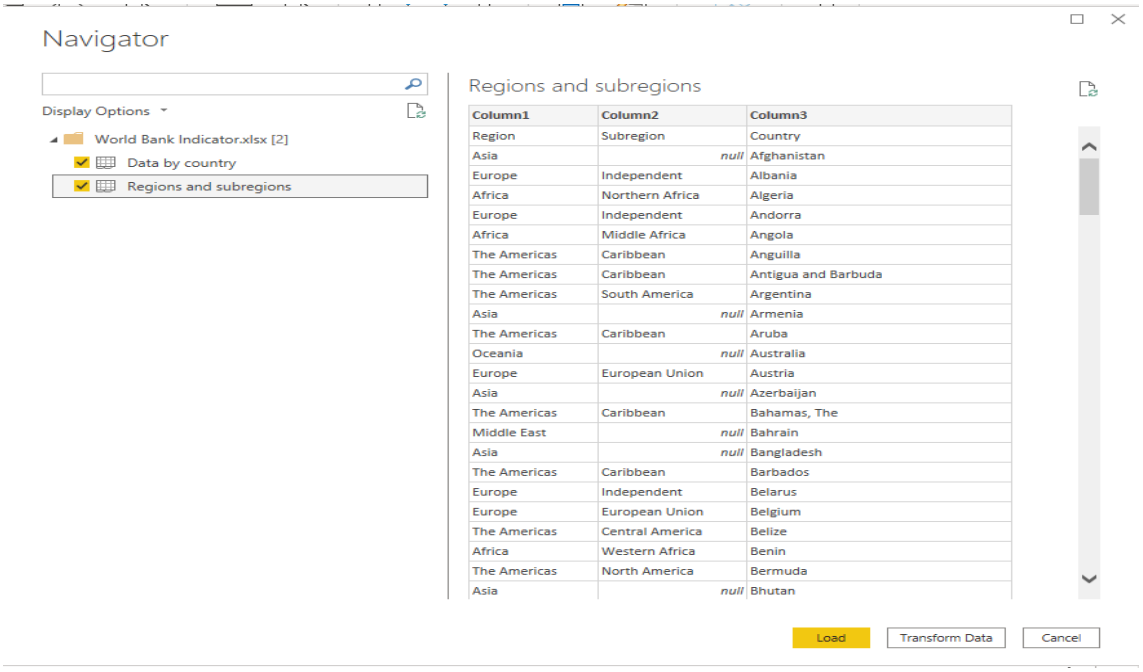
To perform analysis, we need to import data first. Click on '**Get Data**' on the menu bar, we can import data from multiple sources, such as Files (Excel, CSX, PDF, etc.), Databases, Power Platform, Azure, Online Services, and Other (Web, Spark, R script, etc.)



In this project, we will import a dataset from an excel file. The file is "**World Bank Indicator.xlsx**". Click '**Open**' to connect to the dataset.



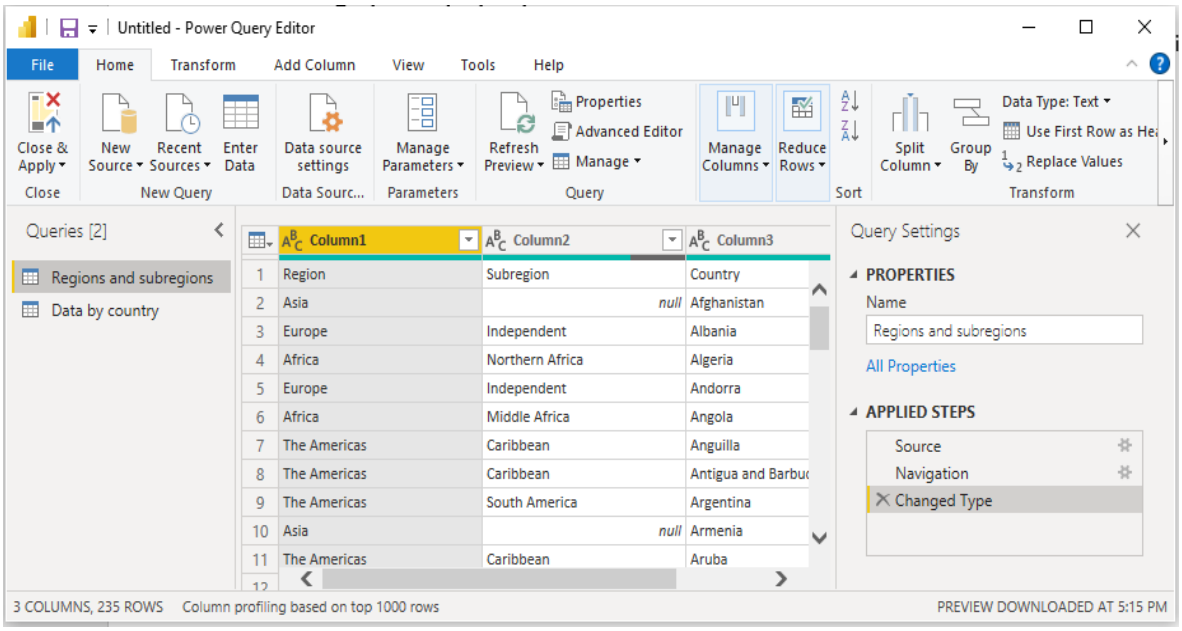
At this point, users can select Load to load the table or Transform data to make changes in the table before users load it.



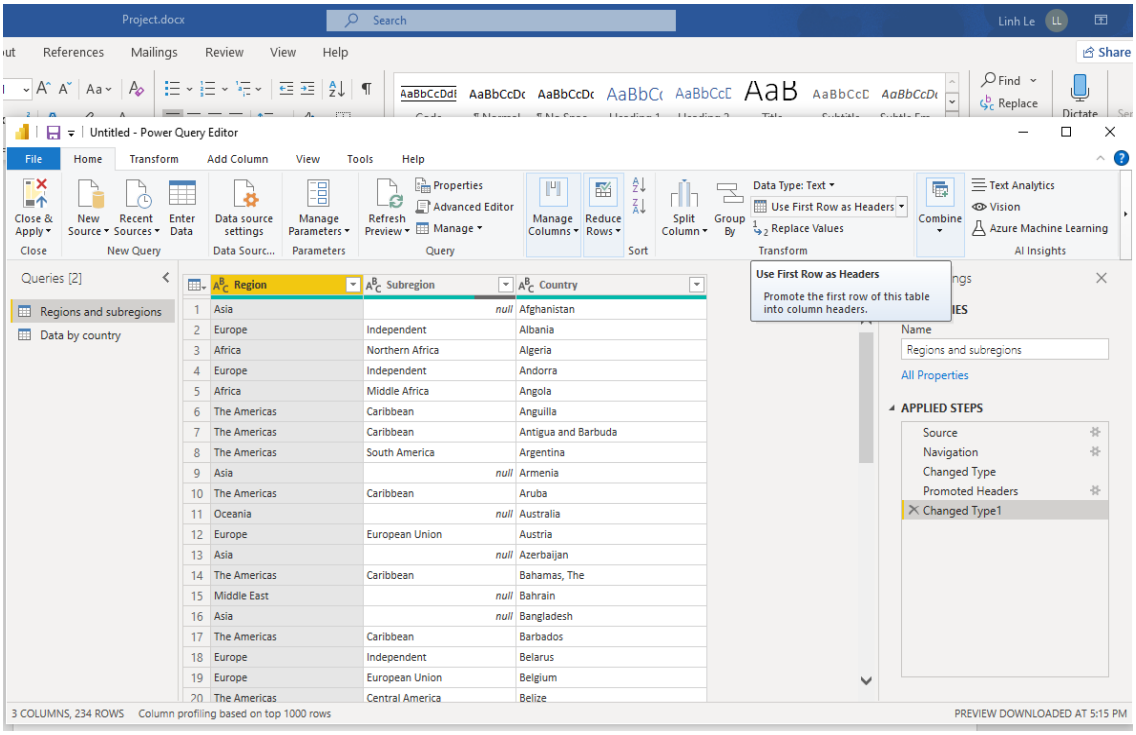
Column1	Column2	Column3
Region	Subregion	Country
Asia	null	Afghanistan
Europe	Independent	Albania
Africa	Northern Africa	Algeria
Europe	Independent	Andorra
Africa	Middle Africa	Angola
The Americas	Caribbean	Anguilla
The Americas	Caribbean	Antigua and Barbuda
The Americas	South America	Argentina
Asia	null	Armenia
The Americas	Caribbean	Aruba
Oceania	null	Australia
Europe	European Union	Austria
Asia	null	Azerbaijan
The Americas	Caribbean	Bahamas, The
Middle East	null	Bahrain
Asia	null	Bangladesh
The Americas	Caribbean	Barbados
Europe	Independent	Belarus
Europe	European Union	Belgium
The Americas	Central America	Belize
Africa	Western Africa	Benin
The Americas	North America	Bermuda
Asia	null	Bhutan

As can be seen from the Navigator, the excel file consists of two sheets: **Data by country** and **Regions and subregions**.

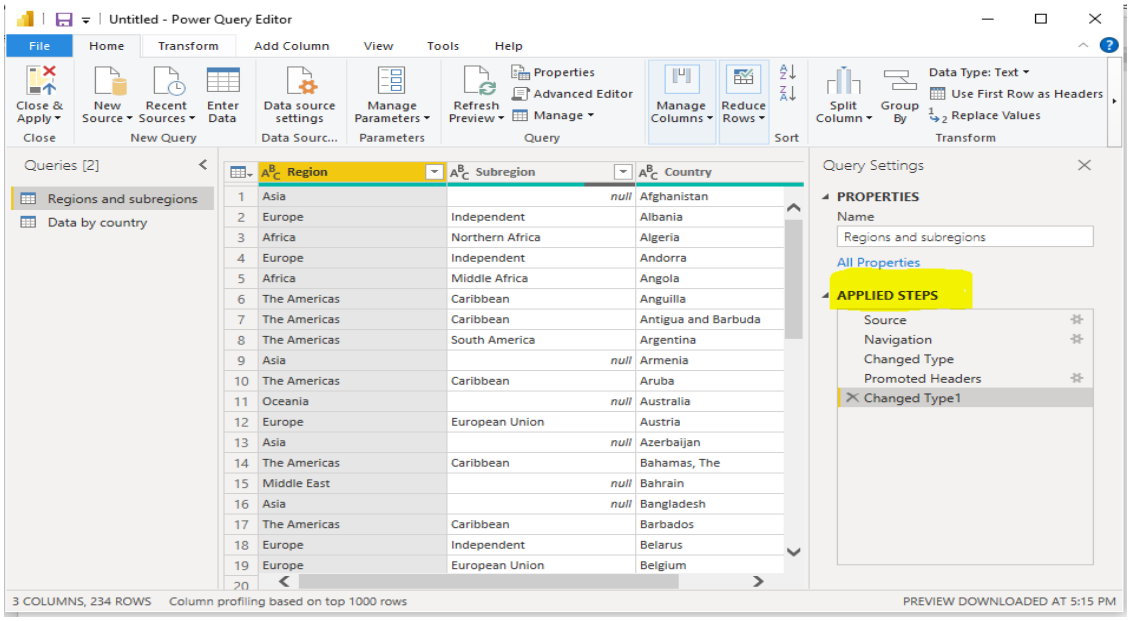
Since the **Regions and subregions** sheet does not have proper column names, so we will click “Transform data” to open “Power Query Editor” to edit data before loading it.



- Click on “**Use First Row as Headers**”, we can have proper column names as below:

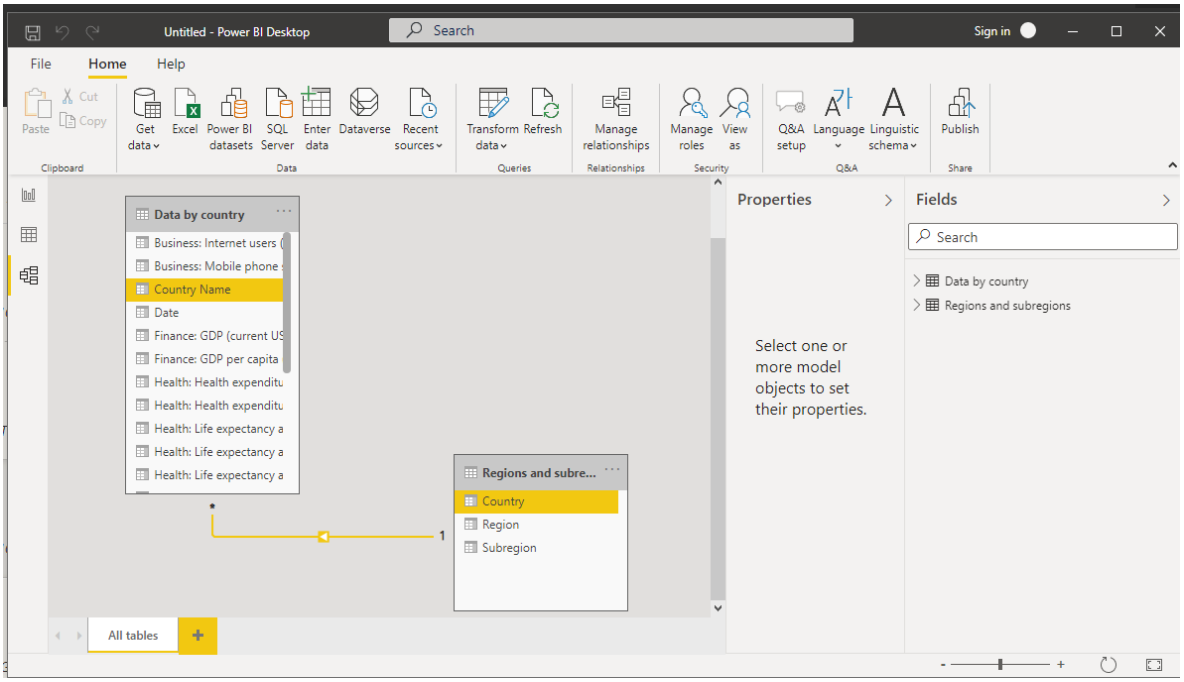


All the steps applied in Power Query Editors are mentioned in the "**Applied Steps**" windows (Highlighted)

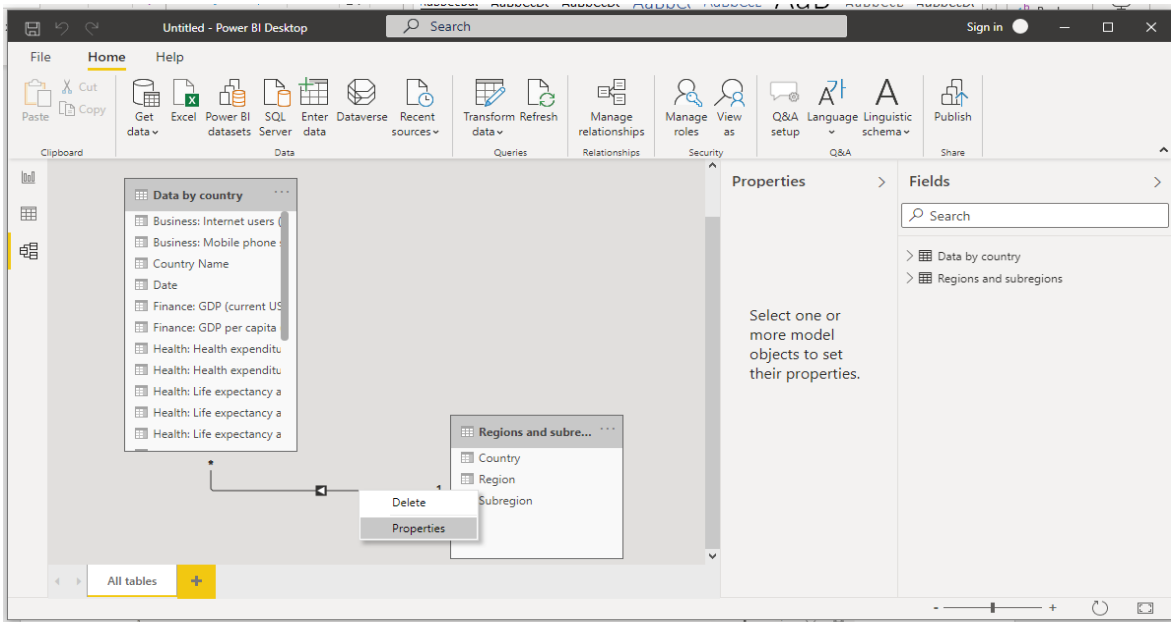


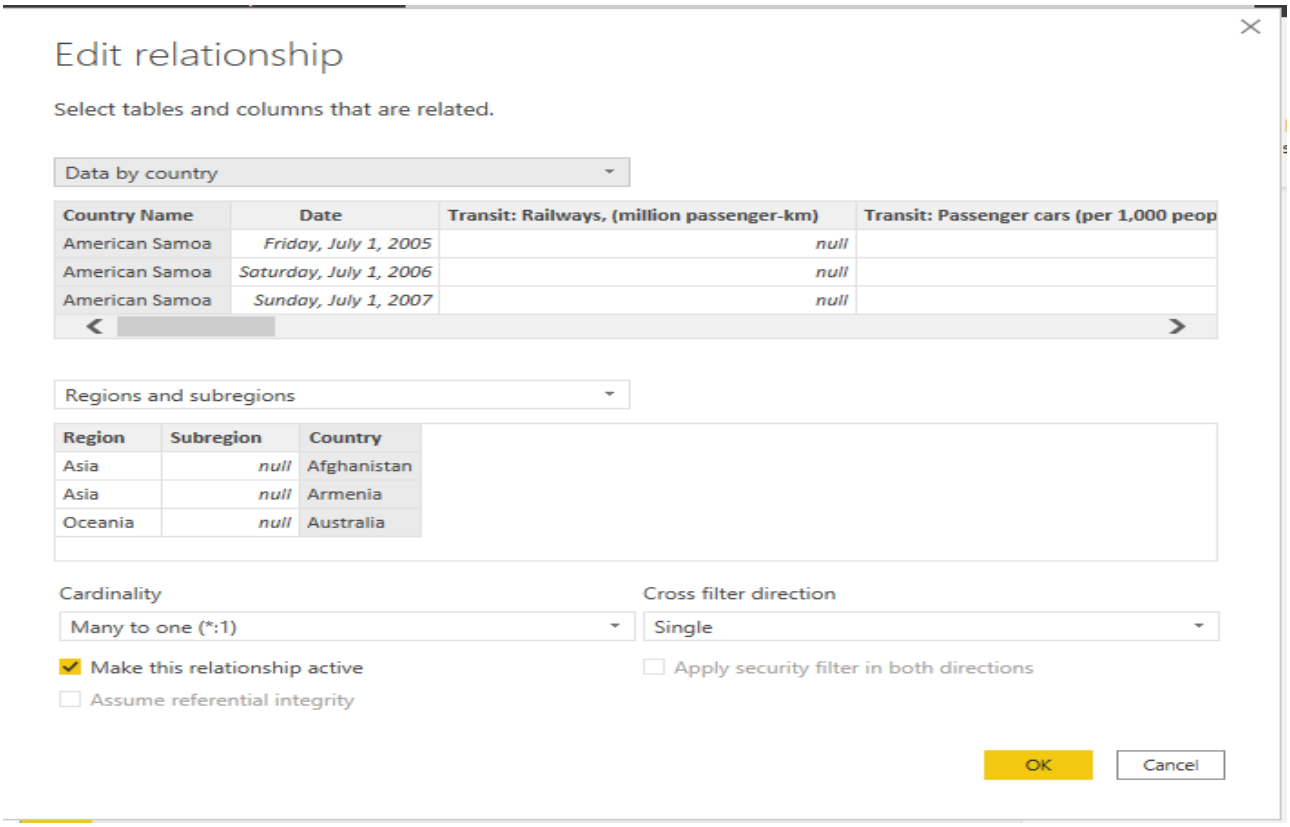
After finish data transformation, hit "**Close and Apply**" on the top-left corner to load the data.

Power BI will automatically create the relationship between 2 sheets



If the automatedly defined relationship is incorrect, users can modify the relationship as desired by right-clicking at the arrow and hitting "**Properties**" to edit the relationship.

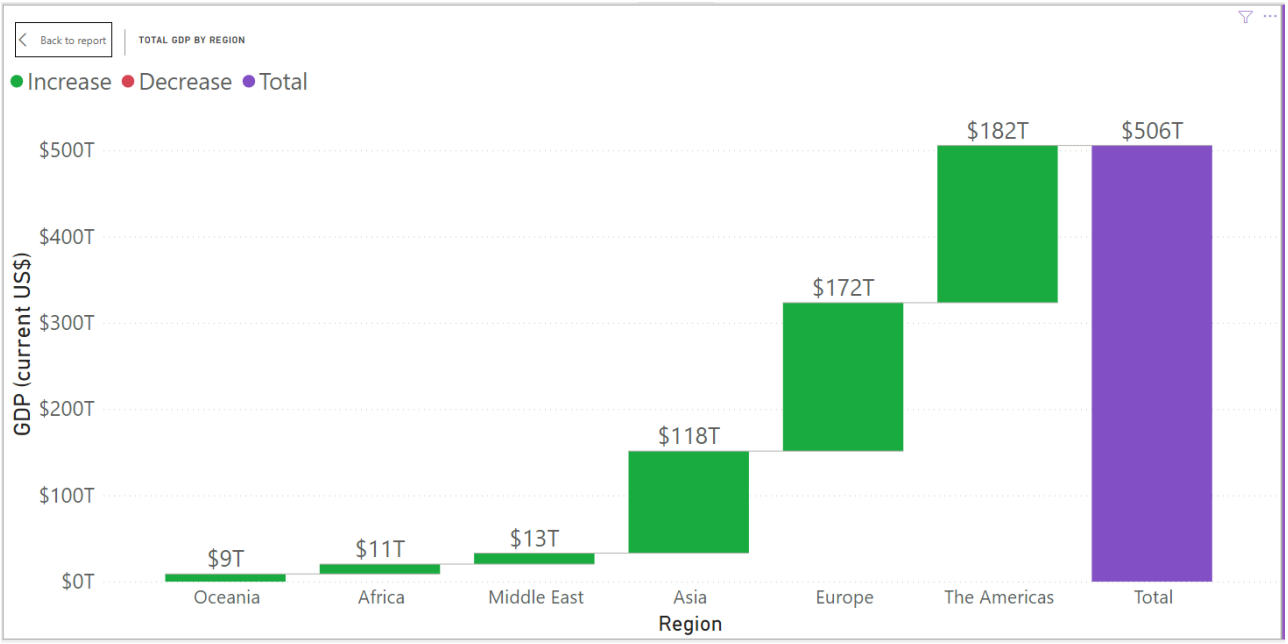




Visualization and Analysis

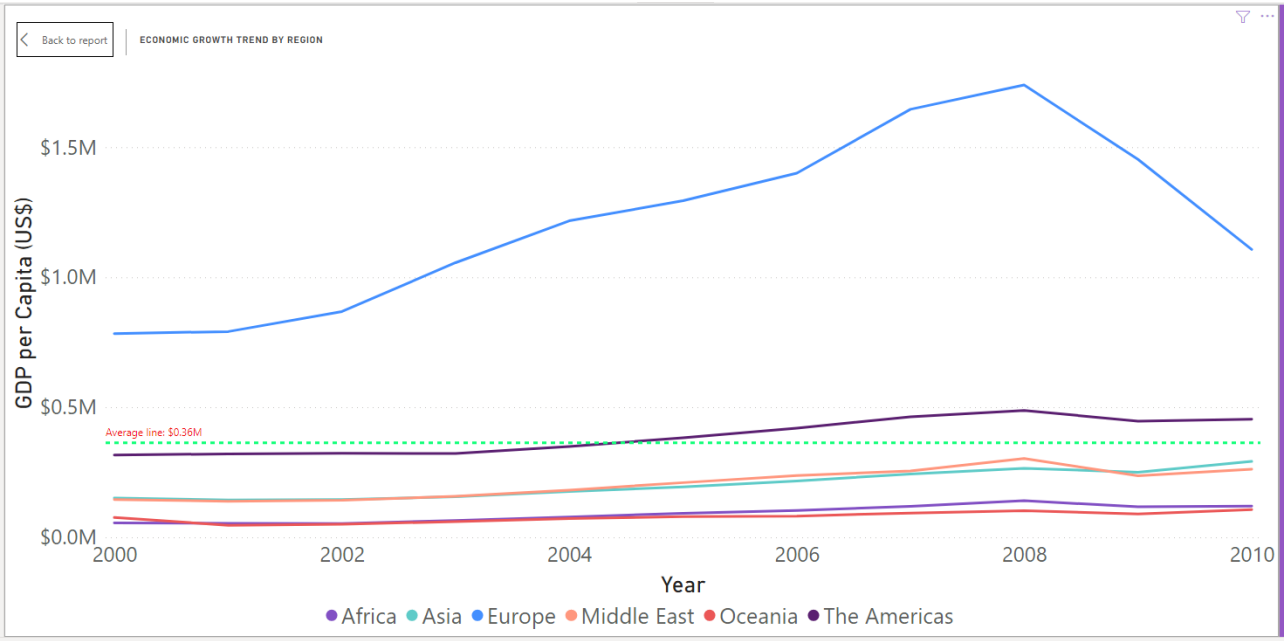
A. Components of the dashboard:

1. TOTAL GDP BY REGION



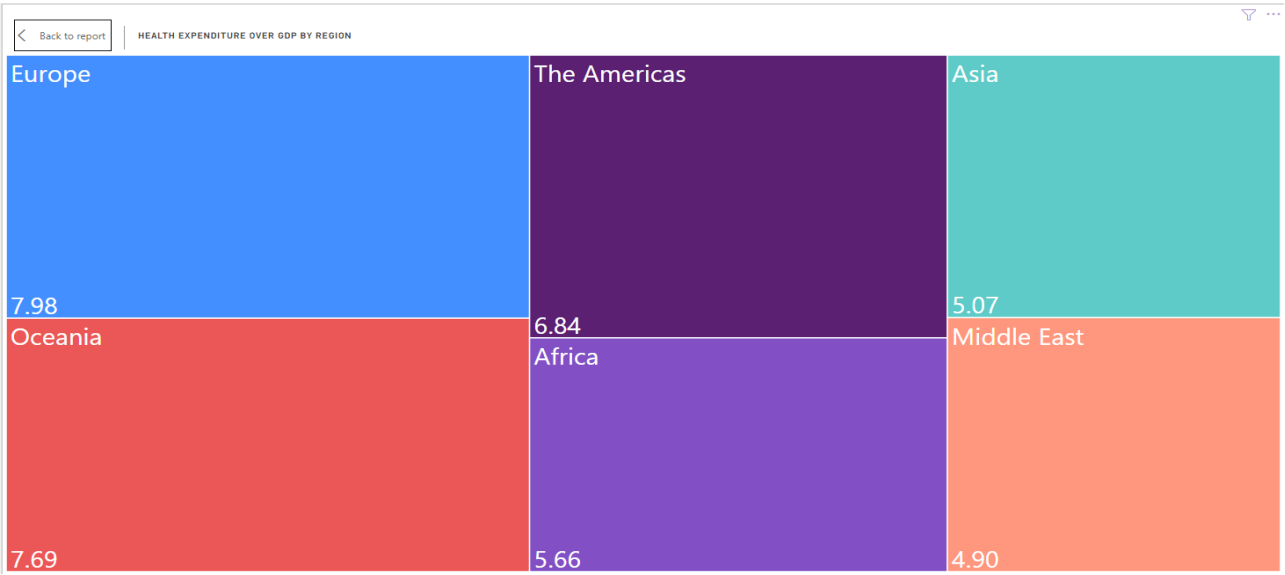
As can be seen from the graph, The Americas achieved the highest total GDP ($182T), followed by Europe with $172T, while Oceania, the region with the lowest GDP amount, attained only $9T in 10 years. There was a big polarization in total GDP between

2. ECONOMIC GROWTH TREND BY REGION



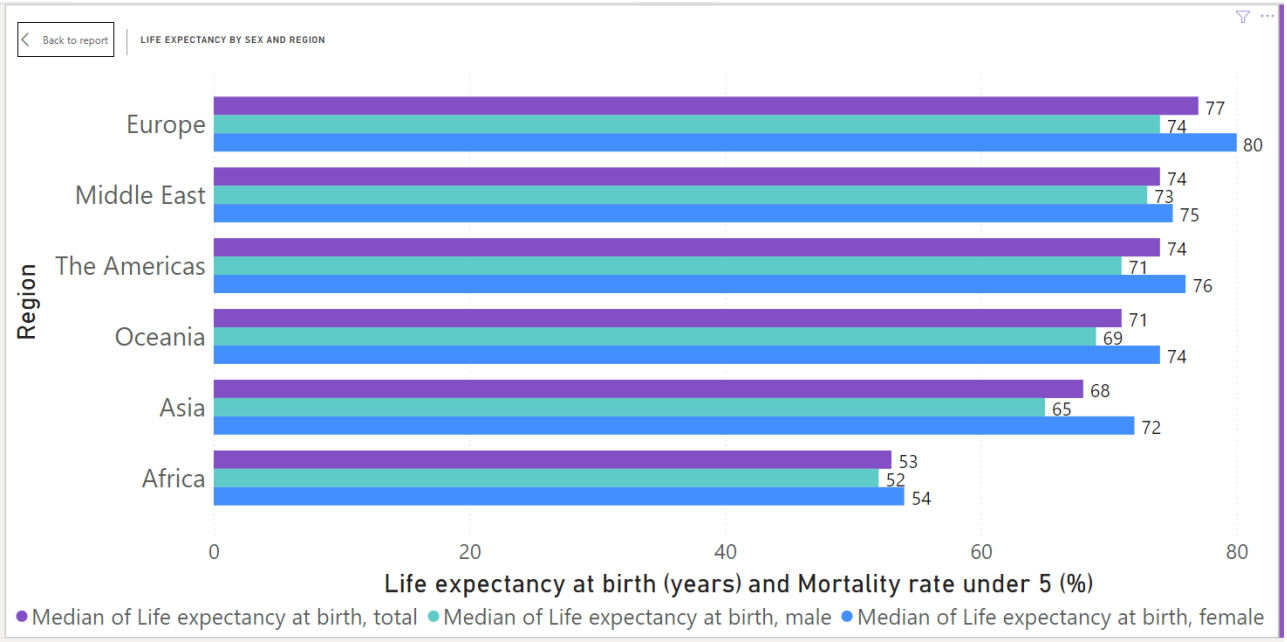
From 2000 to 2005, Europe was the only region having the average GDP per Capita ($0.78M) far above the average line ($0.36M). Since 2005, although the average GDP per Capita of the Americas climbed a little above the average line, the growth rate of Europe witnessed a significant change and left other regions far behind. Its GPD per Capita soared rapidly from 2002 and peaked in 2008 ($1.74M) then plummeted drastically since then.

3. HEALTH EXPENDITURE OVER GDP BY REGION



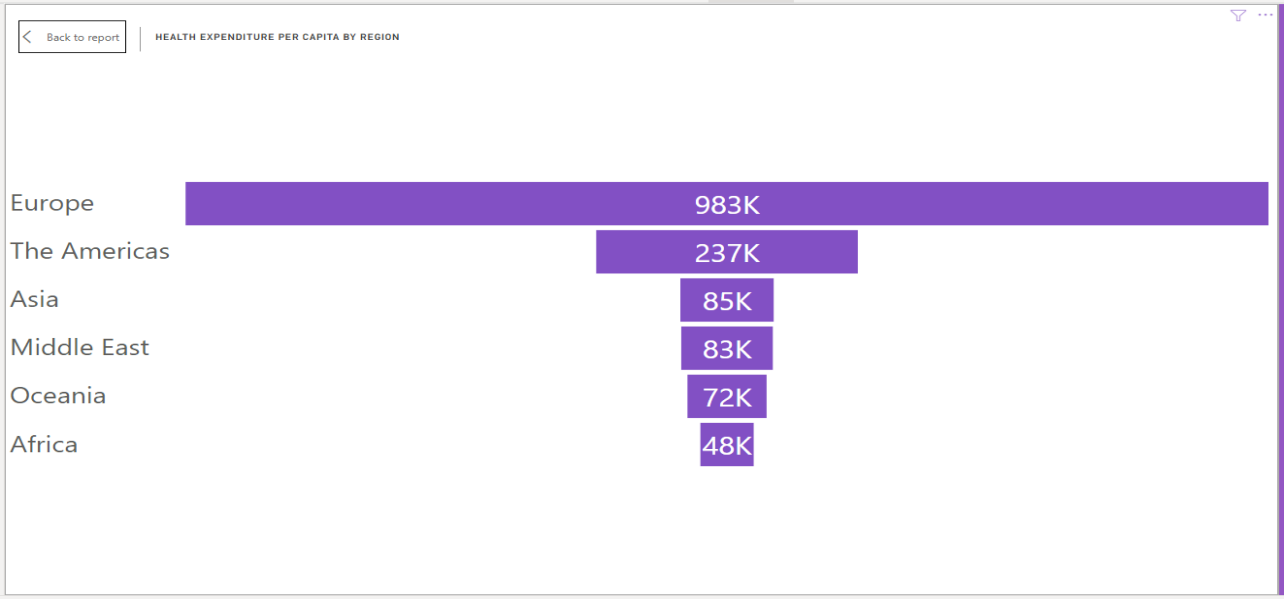
According to the area chart, Europe spent the highest budget (7.98% of total GDP) on the health care system, while Middle East allocated only 4.9% of total GDP for health expenditure and was the region with the lowest spending on health care.

4. **LIFE EXPECTANCY BY SEX AND REGION**



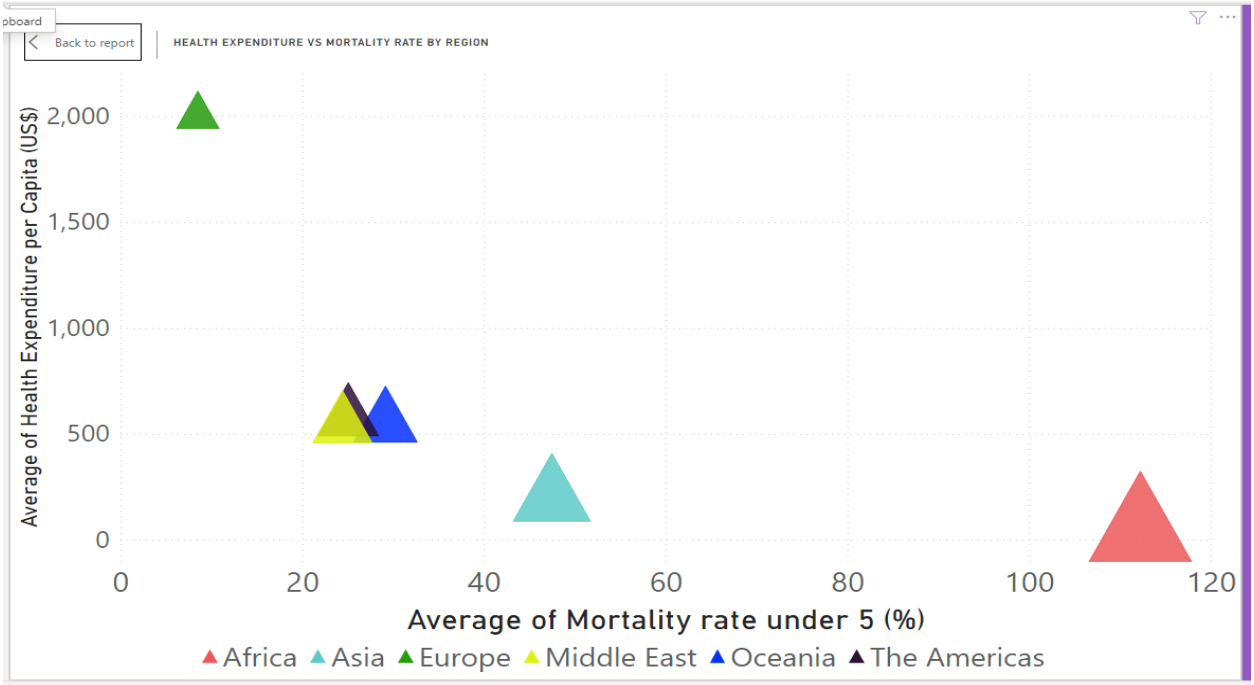
The bar chart illustrates that European people had the highest median of life expectancy (77 years old), while African had the lowest (53 years old) for the total population. In terms of gender, females tended to live longer than males all over the world, and the ranking of regions remained unchanged with respect to sex. European women mostly lived up to 80 years old, while the median of European men's life expectancy was just 74 years old. In Africa, the median life expectancy of men and women did not differ much, 54 vs. 52 years old respectively.

5. HEALTH EXPENDITURE PER CAPITA BY REGION



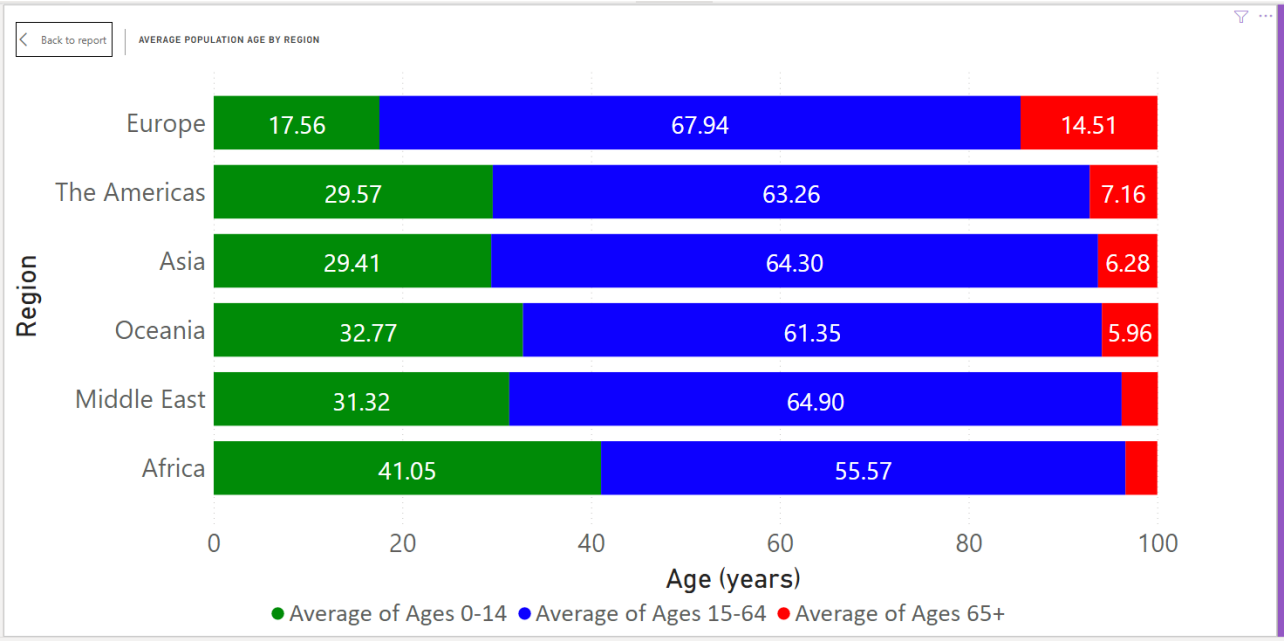
In terms of health expenditure per capita, Europe still kept its leading position with US$983K, while Africa spent the least on health care with only US$48K.

6. HEALTH EXPENDITURE VS MORTALITY RATE BY REGION



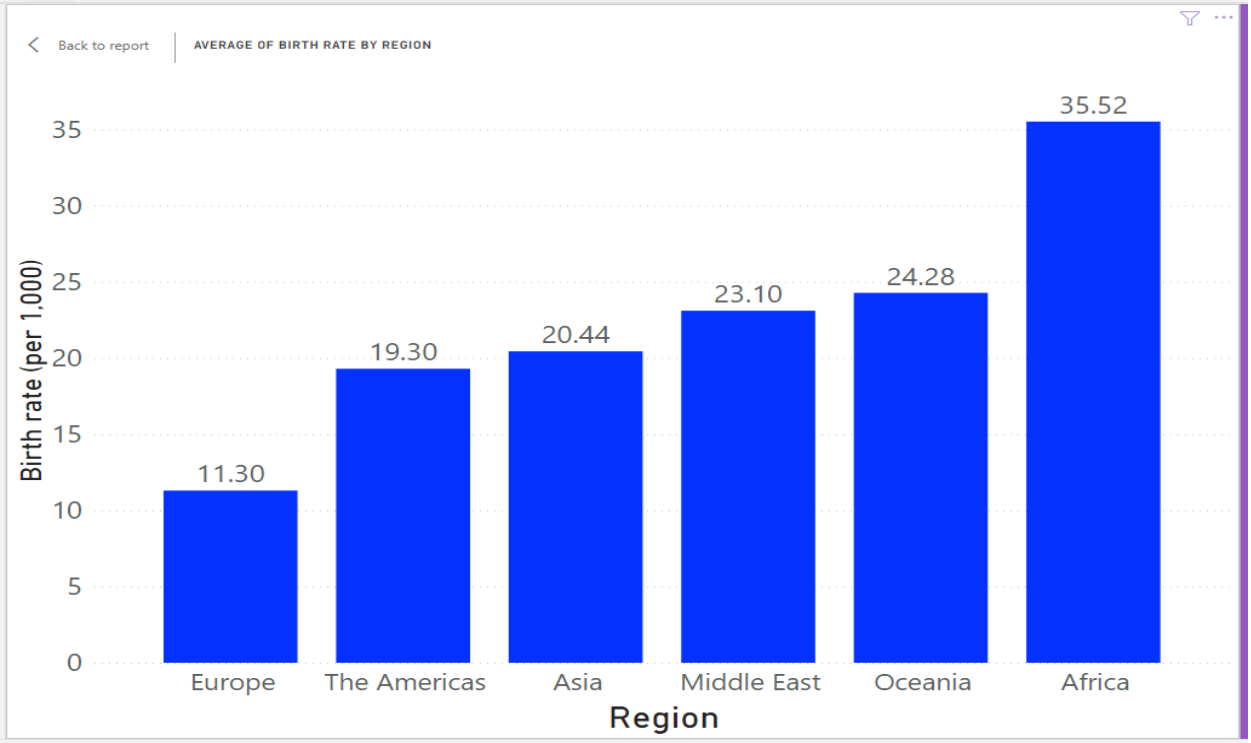
According to the scatter plot, more investment in health expenditure per capita was negatively correlated to mortality rate under 5 years old. Europe had the highest average investment in the health care system (US$ 2,030), as a result, it obtained the lowest average rate of mortality under 5 (8.51 per 1,000 live births). In contrast, Africa suffered the highest average mortality rate under 5 (112.25 per 1,000 live births), as a consequence of the lowest average of health expenditure per capita (US$85.87).

7. AVERAGE POPULATION AGE BY REGION



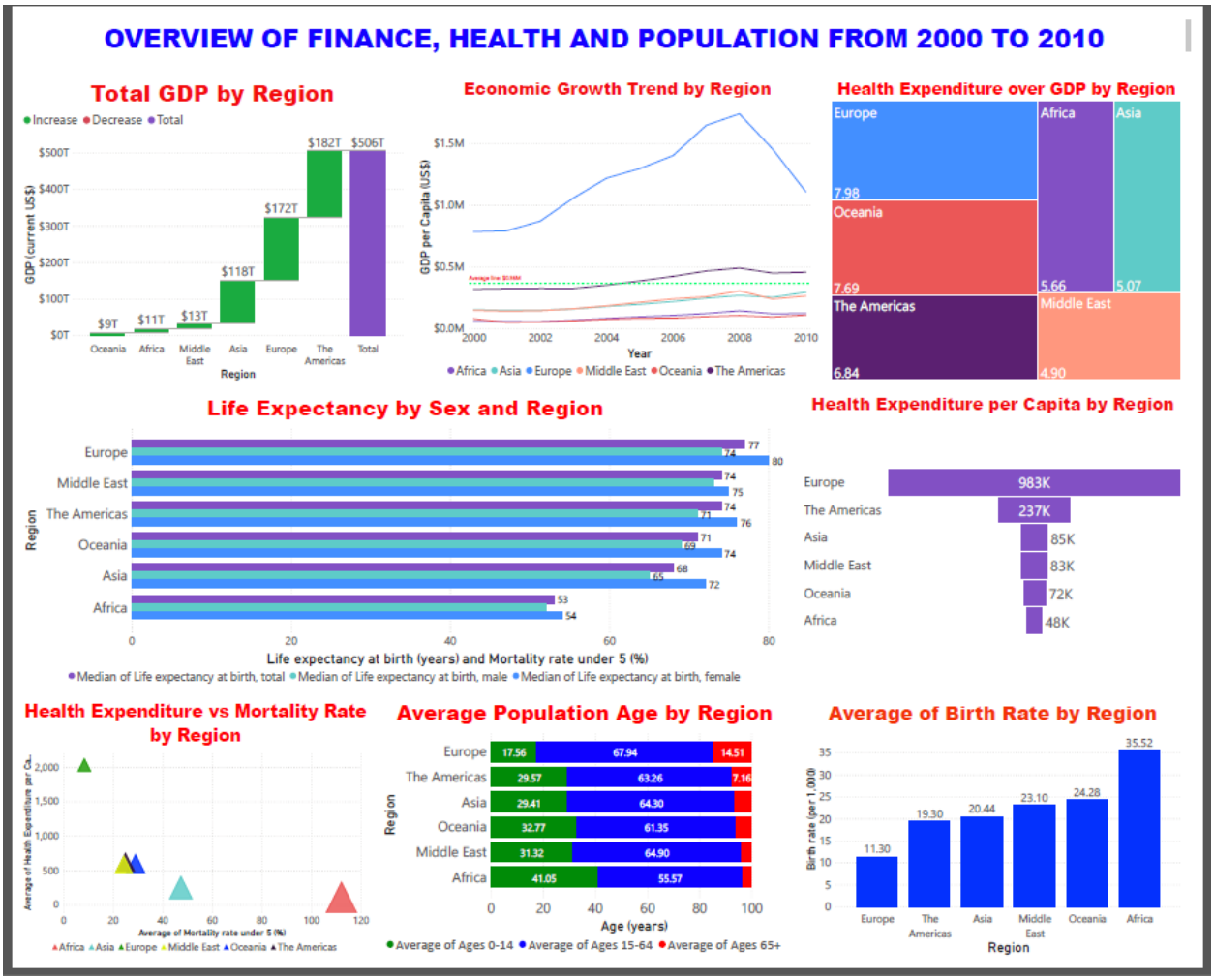
The stacked bar chart demonstrates that Europe had the highest proportion of the population in working age (67.94%), but it was also the region with the lowest percentage of the population less than or equal to 14 years old (17.56%) and the highest portion of the population equal to and over 65 (14.51%), while Africa was in the reverse position for each corresponding age group.

8. AVERAGE BIRTH RATE BY REGION

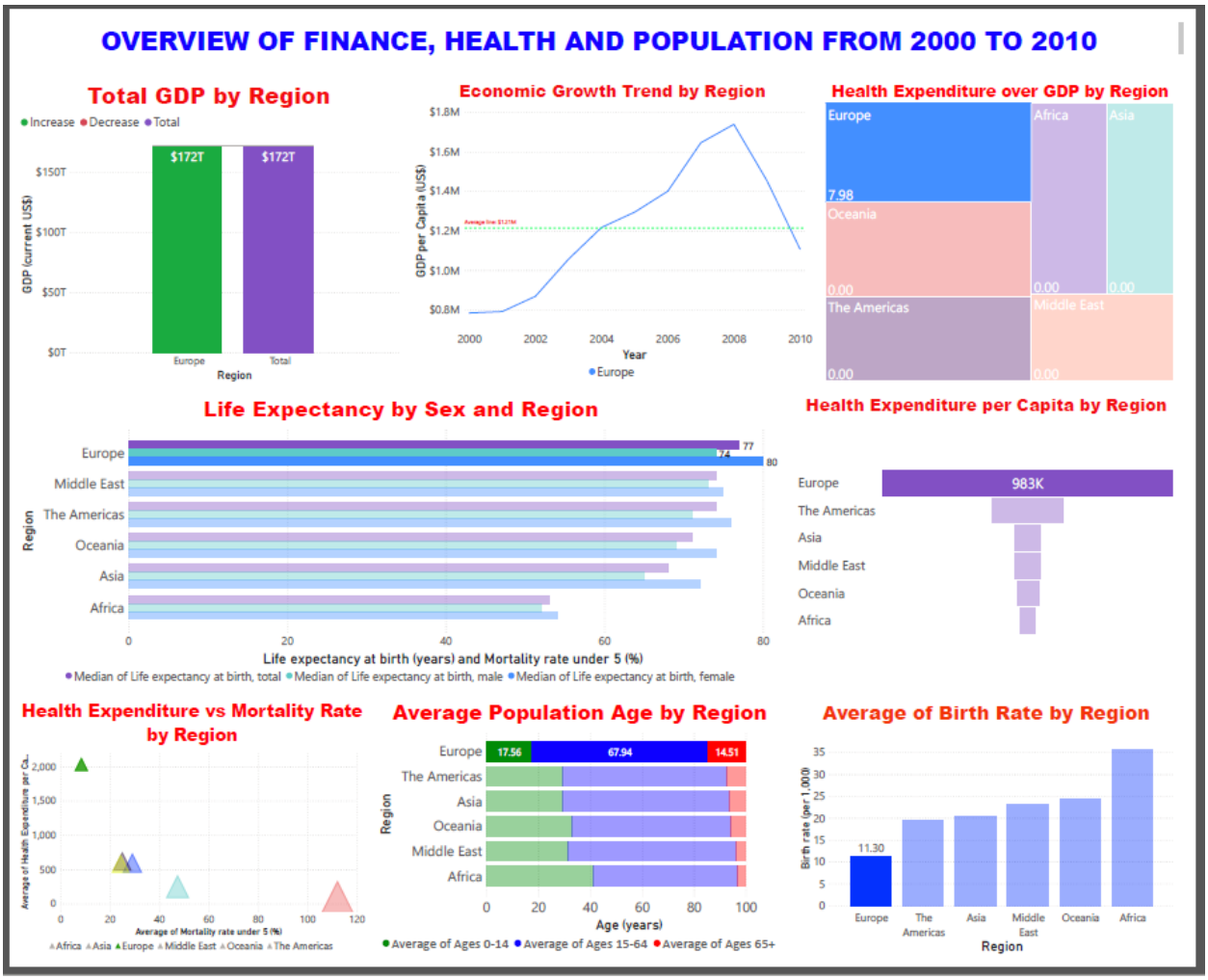


This bar chart explains why Europe had the lowest percentage of people less than or equal to 14 years old since its birth rate was only 11.3 per 1,000, which was the lowest rate among regions. Meanwhile, Africa witnessed the highest birth rate, accounting for 35.52 per 1,000.

B. INTERACTIVE DASHBOARD



C. OVERVIEW OF EUROPE



Results: Overview of Europe from 2000 to 2010

Although the total GDP of Europe was a bit lower than that of the Americas, it gained the highest GDP per Capita. Between 2000 and 2010, while the GDP per capita of the other regions was gradually going up, Europe witnessed a remarkable change in GDP per Capita. Its GPD per Capita soared rapidly from 2000 and peaked in 2008 then plummeted drastically since then. Europe spent the highest budget on the health care system since the health expenditure accounted for approximately 8% of its total GDP. Thanks to huge spending on health care, European people had the highest life expectancy at birth and the lowest mortality compared to the other regions. As can be seen from the graph, Europe was the country coping with the aging population because it has the highest percentage of the population over 65 years old, and the lowest proportion of the population under 14 years old among the regions.

Below are strengths, weaknesses and recommendations:

1. Strengths:
   - Highest GDP per Capita
   - Highest budget on health care
   - Highest life expectancy at birth and lowest mortality under 5
2. Weaknesses:
   - Economic downturn trend
   - Aging population and low birth rate
3. Recommendations:
   - Identify reasons for economic downturns to find appropriate solutions. It could be due to the fact that the labor force became aging, so working efficiency reduced remarkably.
   - Encourage maternity to resolve the aging population issue, enlarge the labor force and boost the growth of the economy

Conclusion

It is proven that Power BI is a great weapon in the visualization and analysis arsenal as it helps users to tell a convincing story underlying the data as well as communicate that story to their audience effectively. The audience could have a quick understanding through the interactive dashboards about overview of finance, health, and population of different regions in the world from 2000 to 2010. Among the regions, Europe prominently achieved the highest GDP per Capita, spent the highest budget on the health care system, as a result, gained the highest life expectancy at birth and lowest mortality under 5, though it coped up with some difficulties such as economic downturn trend, aging population, and low birth rate. Some suggestions could be drawn for the existing problems, for example, the authorities should identify reasons for economic downturns to apply appropriate solutions, and encourage people to give more births to resolve the aging population and shortage of labor force issue to foster economic growth.

REFERENCES

Iseminger D., & Sharkey K., (2019). Tutorial: Shape and Combine Data in Power BI Desktop. *Microsoft*. Retrieved from https://docs.microsoft.com/en-us/power-bi/connect-data/desktop-shape-and-combine-data

Myers P., Follis K., Sharkey K., & Blythe M., (2019). Data Reduction Techniques for Import Modeling. *Microsoft*. Retrieved from https://docs.microsoft.com/en-us/power-bi/guidance/import-modeling-data-reduction

Ouaknine E., (2020). Power BI vs Tableau: 10 Questions to Ask Before Choosing the Appropriate BI Tool. *Upslide*. Retrieved from https://www.upslide.net/en/power-bi-vs-tableau/

Power BI Features. *Data Flair*. Retrieved from https://data-flair.training/blogs/power-bi-features/

Tableau vs Power BI: Key Differences. *Guru99*. Retrieved from https://www.guru99.com/tableau-vs-power-bi-difference.html