

Datathon - 6

¹Arjun Verma : IMT2017008

CS732/DS732: Data Visualization

Course Instructors : Prof. Jaya Sreevalsan Nair
Technical Report - 6

***Abstract.** This technical report contains an overview of the methodology involved in generating the visualizations of the provided [dataset](#) consisting of an overview of countries. This report is in continuation to Datathon 5 and attempts to generate map-based visualizations to build upon the inferences from the previous report.*

1. Introduction

The provided dataset contains information about different statistical measures that can be used to track the development status of the countries for which these variables have been provided. These include a number of variables such as GDP per capita, life expectancy of men and women at various ages, economic activity, unemployment rates, etc. *In this assignment, we try to continue on the detailed case study of the economic development of the given countries.*

2. Tools and Methodology

2.1. Tools

The exhaustive set of libraries used for generating the final inference visualizations involve :

- numpy
- pandas
- plotly
- geopandas
- matplotlib
- PIL

Note: The aforementioned tools are used solely for the generation of choropleths. The generation of heatmaps is not suitable for the dataset provided as the data is limited to countries and heatmaps are generally used to visualize information which do not follow well defined political boundaries.

2.2. Methodology

In this section, we provide the distinct characteristics involved in generating choropleths and the reasonings behind them. However, before we dive into the individual techniques, let's have a clarity on the basic workflow of the implementation.

2.2.1. General Methodology

For getting to the final combined inference, the dataset has to go through two stages. The workflow proceeds in this order :

- First, the dataset is collected and passed along to `economic_analysis.py`. These generated a spatial visualization for each time frame.
- These time frames are then passed onto `gif_generator.py` to output the final animation for the spatio-temporal analysis.

A general point to note with respect to the dataset is that the year 2016, has absolutely no values for the essential features across all countries and hence has been specifically truncated from the dataset. The latest trends have been analyzed upto 2015.

2.2.2. Methodology for Map-based visualizations

Distinct techniques (characteristics) for gathering inferences :

- In total of four different perspectives have been generated using choropleths. The perspective of utmost importance is the combined analysis of population and GDP per capita. Intermediate perspectives focusing solely on GDP per capita, import goods and export goods form the other three visualizations.
- For the combined analysis, categories have been made based on two thresholds: **the average population of the dataset** and **the GDP per capita for developed countries**. Four distinct colors have been chosen to display the following four permutations and perform a combined analysis :
- For the sole analysis of **GDP**, a **diverging colormap (RdBu)** with it's midpoint fixed at the threshold for a country to be labelled as **developed** has been used. This helps us notify the year when a country has had it's development status changed.
- For the sole analysis of **Import and Export**, a **sequential colormap (Plasma)** has been used to keep a track of them for the countries over the given timeframe.
- A combined analysis from the sole visualizations has also been made.
- A study using pie chart visualizations to focus on sectors contributing the GDP has also been made.

2.3. Sample visualizations

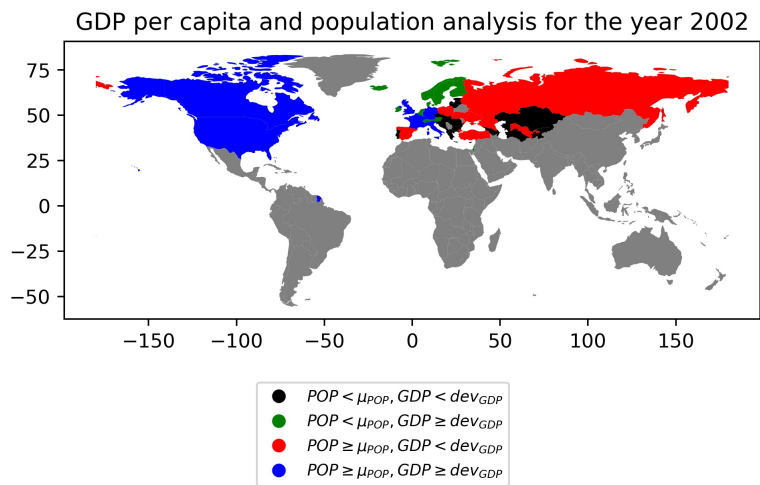


Figure 1. Sample of the final combined visualization for the year 2002

GDP per capita of the countries in dataset for the year 2002

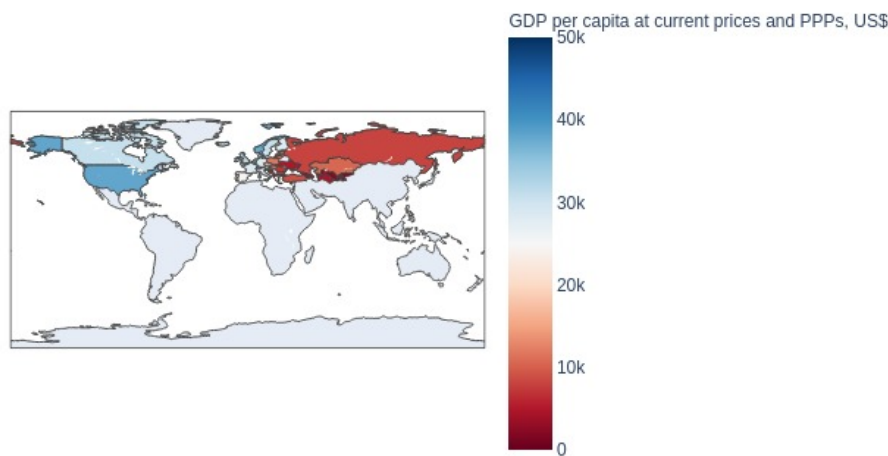


Figure 2. Sample GDP per capita intermediate visualization for the year 2002

Import of goods percent of GDP of the countries in dataset for the year 2002

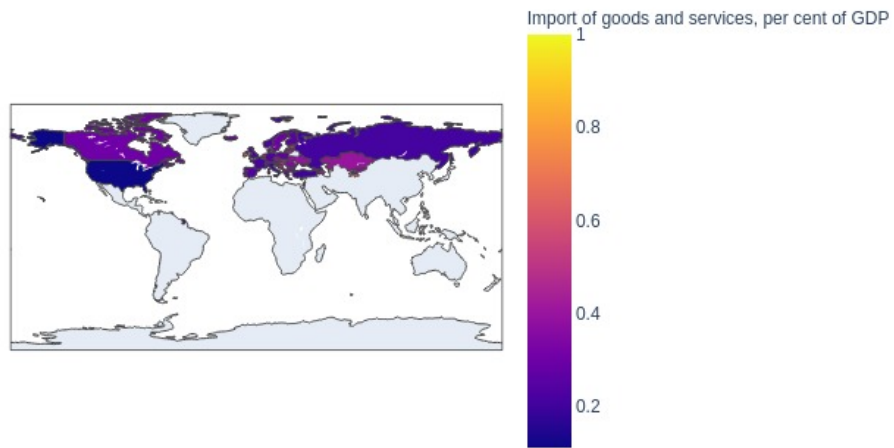


Figure 3. Sample imported goods intermediate visualization for the year 2002

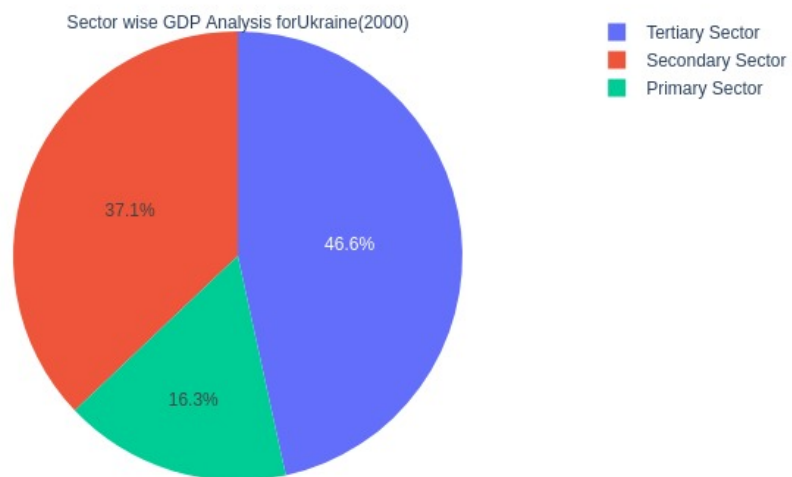


Figure 4. Sample sector wise GDP analysis for Ukraine (2000)

3. Inferences

1. ***Choropleth for the combined analysis of population and GDP :*** In our combined analysis of **GDP per capita** and **Total population**, we try to see how does a GDP change effect the total number of people in the countries in the given dataset. The countries for whom the data is not available have been greyed out. If more number of people belong to a developed country, the better the facilities for each person. Similarly, in case of a GDP drop it is important to see the effect it would have on the number of people and how their daily facilities might get displaced. Thus, this visualization can not only analyze the GDP per capita and total population individually but can also make combined inferences.

1. We can see the color change for particular countries and infer in which year they reached the developed status. Some of the notable countries that change colors because of the change in development status are **Spain, Kazakhstan, Cyprus** and others.
2. There are only two color changes taking place in the visualization : **red to blue** or **black to green**. This indicates that there hasn't ever been a mass migration of sorts in the given countries for the given time frame. This also shows that the GDP per capita increased because of the economic activities and not a decrease in population.

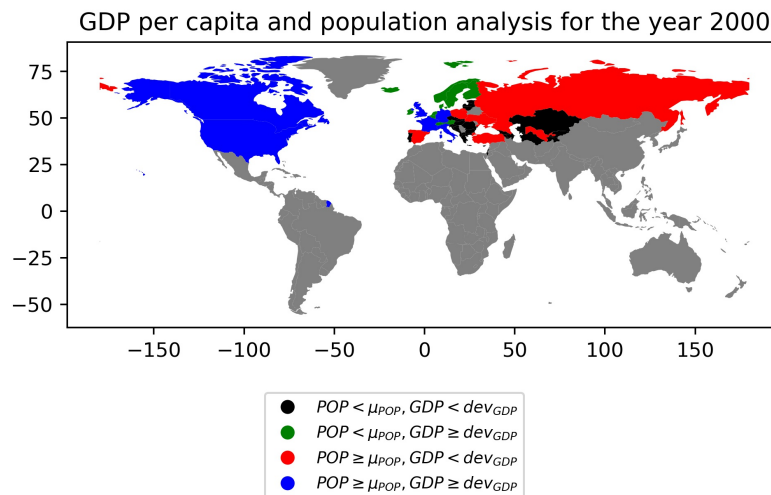


Figure 5. The combined analysis at the beginning of given time frame (2000)

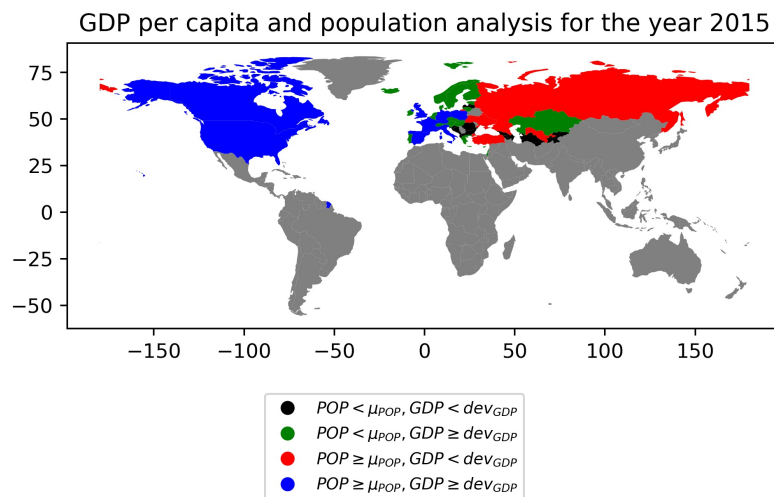


Figure 6. The combined analysis at the end of given time frame (2015)

Color Choice	Developed	Developing
Above avg pop	Blue	Red
Below avg pop	Green	Black

Figure 7. Chosen color scheme for the visualization

2. **Individual GDP analysis using choropleths :** Before we relate the various individual choropleth analysis, let's focus on the most important of the lot : **The GDP Analysis.** We use a diverging colormap to show the distinction between developed and developing countries and study the entities separately in the same visualization. The transition point from developing to developed has been chosen as the mid point. The inferences made are as follows :

1. We can see that the colors for developed countries keep becoming lighter in the period **2007-2009**. This is an indication of the **global economic crisis** that hit the world due to the housing bubble collapse in the United States.
2. **Russia** is the only renowned country that was not hit as hard the rest of the countries. This is evident from the lighter shades of color across the aforementioned time frame. This is largely because Russian macroeconomics is largely decoupled from the rest of the world and they do not follow the traditional European models.
3. The shift in development status for countries like **Spain** and **Kazakhstan** can also be inferred from this animation.

GDP per capita of the countries in dataset for the year 2006

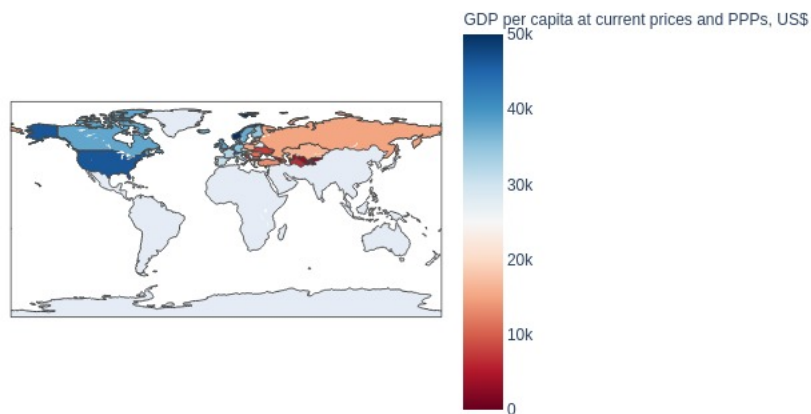


Figure 8. 2006 : Countries before the global economic crisis

GDP per capita of the countries in dataset for the year 2009

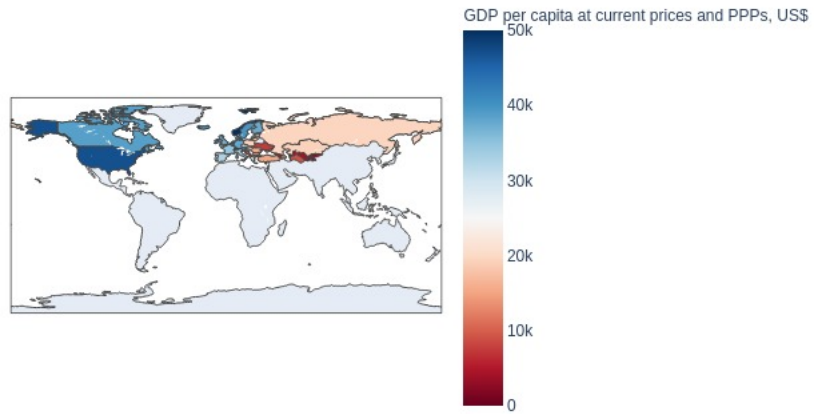


Figure 9. 2009 : Countries after the global economic crisis

GDP per capita of the countries in dataset for the year 2000

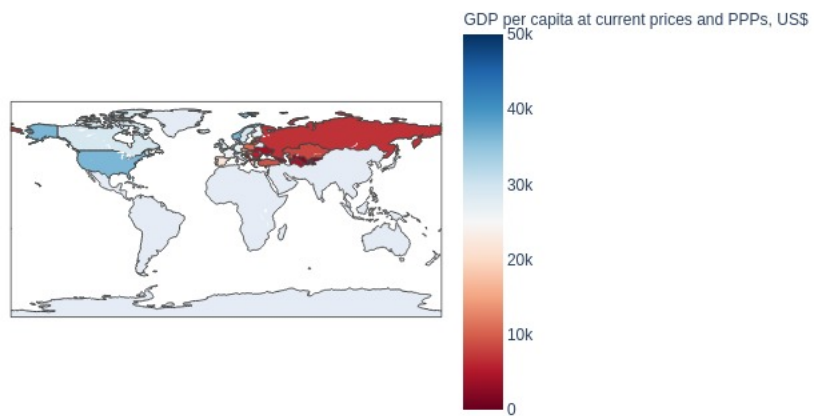


Figure 10. Countries at the start of the given time frame in 2000

GDP per capita of the countries in dataset for the year 2015

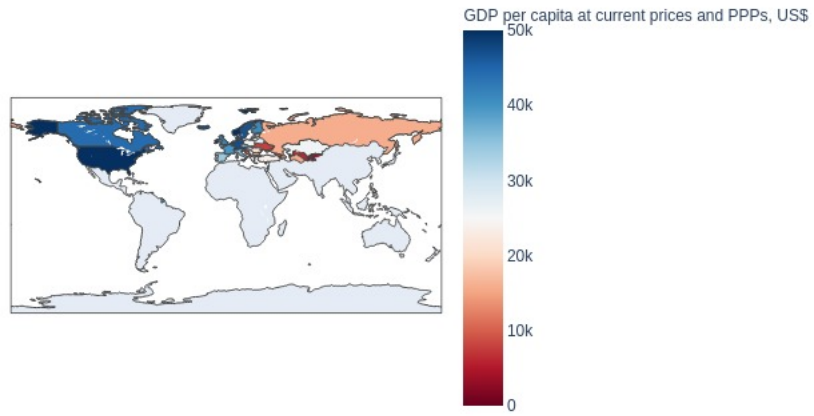


Figure 11. Countries at the end of the given time frame in 2015

3. Unemployment rate analysis Let us now look at one of the major consequences of the **global economic crisis** :

1. Due to the global economic crisis of 2007-2009, there was an increase in unemployment rates all across the world, which had everlasting consequences for some of the countries. Before and after this crisis, the general trend visible was of decreasing unemployment rates. **Spain** was the most effected by this and it has seen an ever since increase in unemployment rates as a result of the crisis. It can again be reiterated that **Russia** dealt with this crisis the best.

Unemployment rates for the countries in dataset for the year 2006

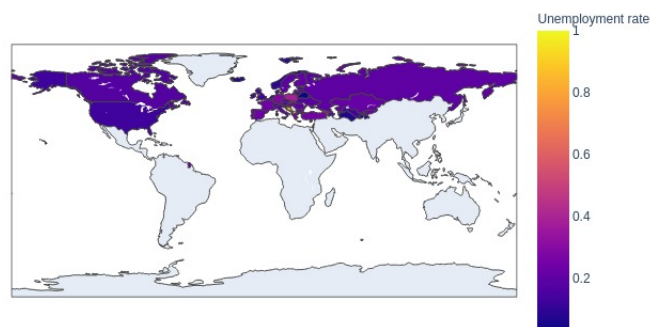


Figure 12. 2006 : Countries before the global economic crisis

Unemployment rates for the countries in dataset for the year 2009

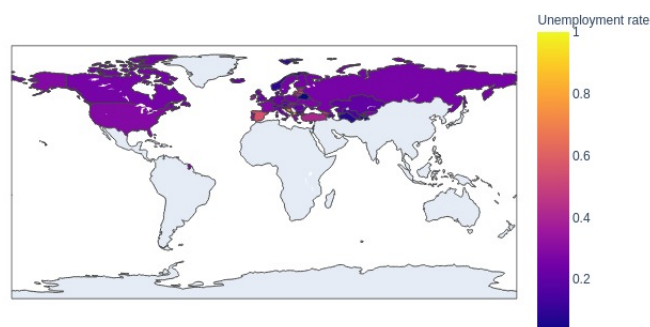


Figure 13. 2009 : Countries after the global economic crisis

Unemployment rates for the countries in dataset for the year 2000

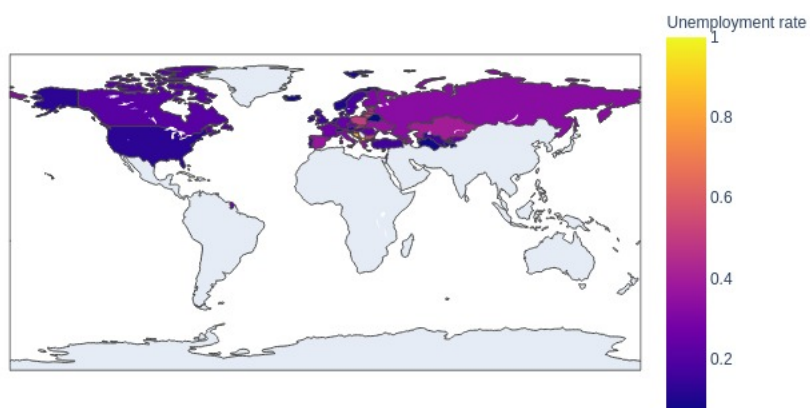


Figure 14. Countries at the start of the given time frame in 2000

Unemployment rates for the countries in dataset for the year 2015

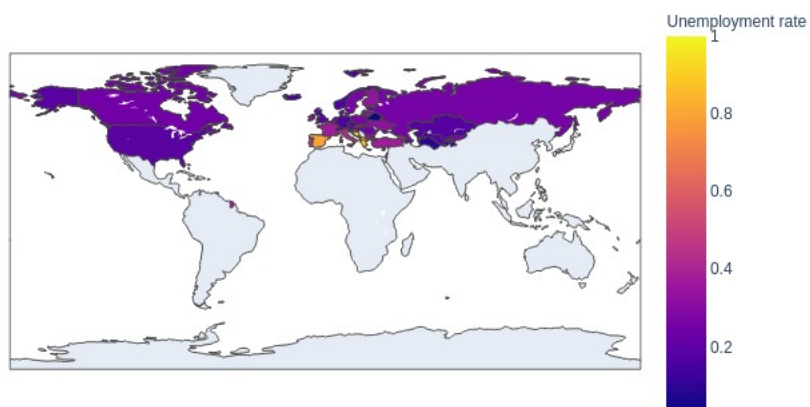


Figure 15. Countries at the end of the given time frame in 2015

4. ***Import and Export Analysis :*** In this section we would highlight how the importing and exporting of final goods correlates to GDP :

1. We can see a general trend that countries which have shown a growth in GDP have shown relatively less dependency on imports for the growth years.
2. The lesser dependency on imports does not necessarily mean an increase in exports. However, these trends were visible in **Kazakhstan** and **Cyprus** to name a few.

Import of goods percent of GDP of the countries in dataset for the year 2000

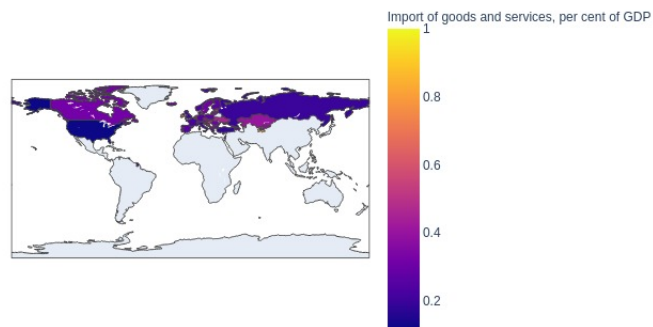


Figure 16. Importing dependencies at the beginning of time frame (2000)

Import of goods percent of GDP of the countries in dataset for the year 2015

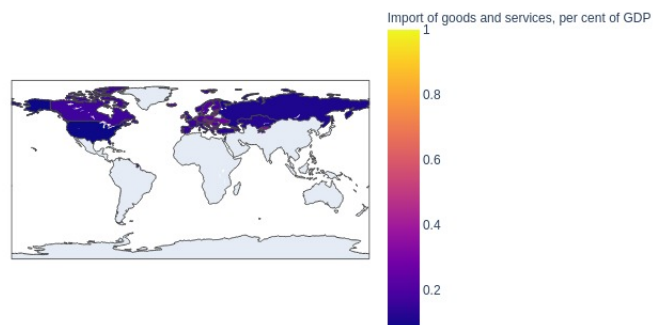


Figure 17. Importing dependencies at the end of time frame (2015)

Export of goods percent of GDP of the countries in dataset for the year 2000

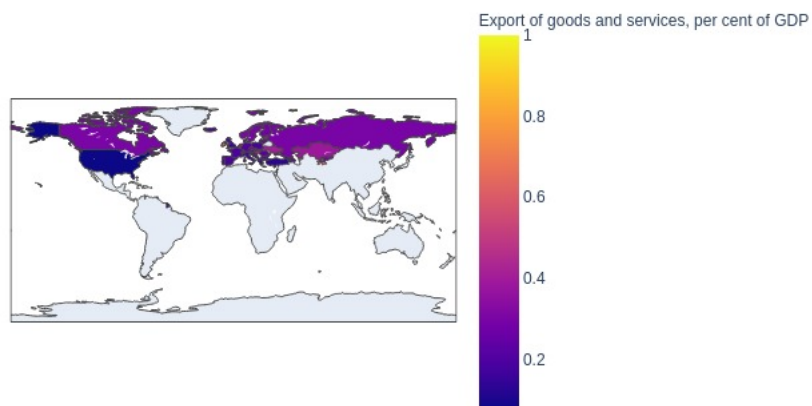


Figure 18. Export share at the start of the given time frame in 2000

Export of goods percent of GDP of the countries in dataset for the year 2015

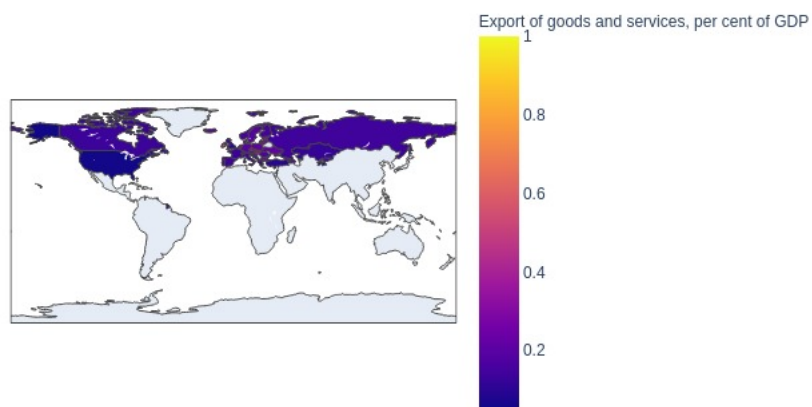


Figure 19. Export share at the end of the given time frame in 2015

5. ***Sector wise analysis of GDP :*** In this section, we analyzed the different sectors that contribute to the GDP of a country. A certain degree of feature engineering was required for this section of analysis. We have resorted to animating pie charts for this analysis. The GDP of a sector is traditionally contributed to by three sectors : **Primary**, **Secondary** and **Tertiary**. The **Primary Sector** is essentially the raw material manufacturing sector and involves agriculture, forestry, fishing, etc. In our dataset, we only have the contributions from agriculture sector and thus have resorted to those as our primary sector. The **Secondary Sector** traditionally involves industries which produce finished materials. From our dataset, the industry and energy contributions to GDP were combined under the secondary sector. The **Tertiary Sector** is the services providing sector. For our analysis, trade, hospitality, finance services, public administration, health, other services all have been combined under one hood of the tertiary sector.

1. The inference of utmost importance from these visualizations is that as and when a country becomes more developed, it is the **Tertiary Sector** which gets expanded and supported the most. Thus, governments should support the **Tertiary Sector** to bring about a growth in GDP in their countries.
2. Supporting the above argument, both **Primary** and **Secondary** sectors undergo huge drops as and when a countries progresses. The traditional trend that is visible is that the **Tertiary Sector** contributes the most, then the **Secondary** and finally the **Primary**.
3. One example each from the developed country pool (**Luxembourg**) and developing country pool (**Ukraine**) have been picked and represented below. The rest of the visualizations can be seen in the Outputs folder submitted along with this document.

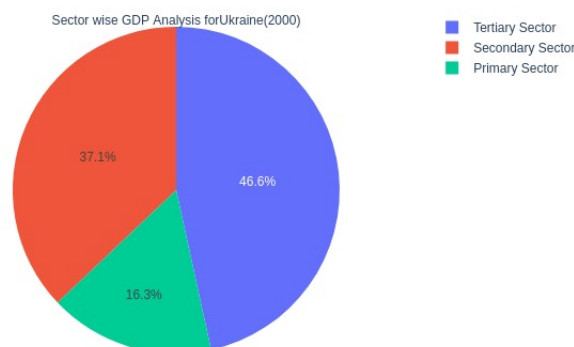


Figure 20. Ukraine's sector wise distribution at the beginning of time frame in 2000

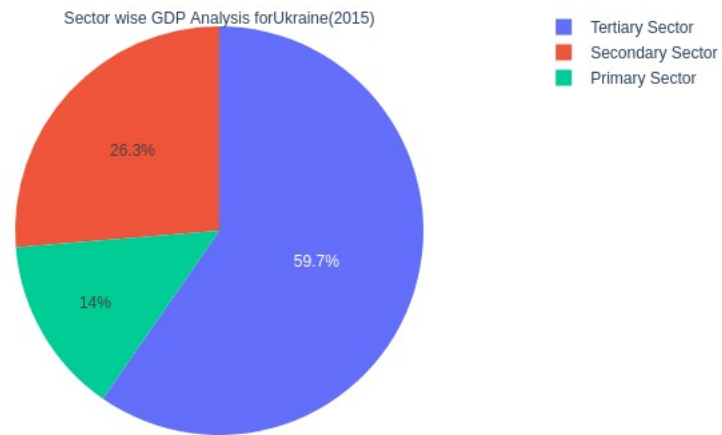


Figure 21. Ukraine's sector wise distribution at the end of time frame in 2015

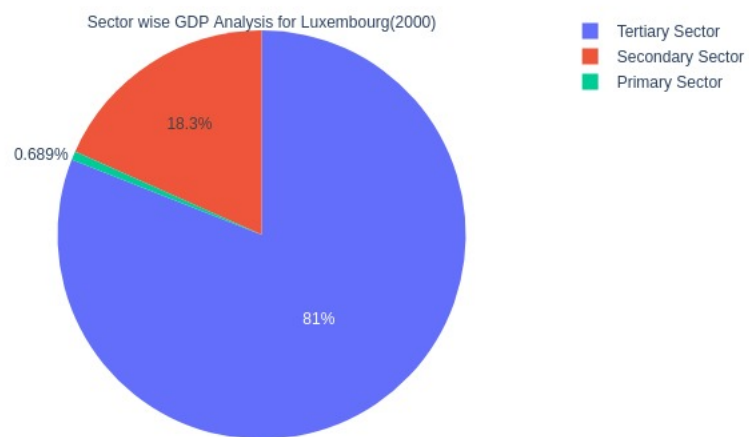


Figure 22. Luxembourg's sector wise distribution at the beginning of time frame in 2000

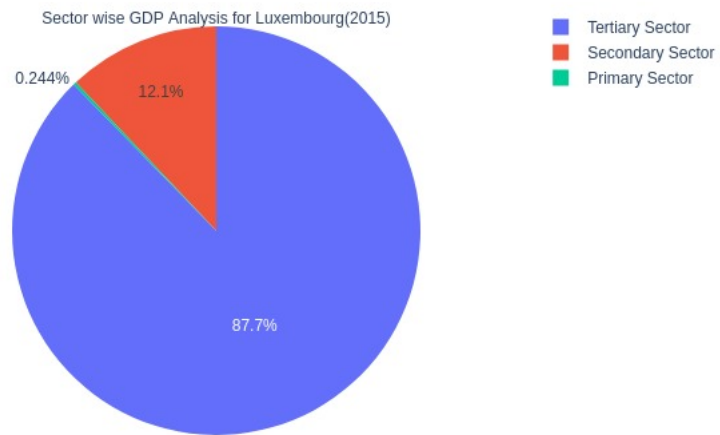


Figure 23. Luxembourg's sector wise distribution at the end of time frame in 2015

4. Conclusion

In conclusion, it is worth to mention that the report has successfully elaborated on all the deliverables. The tools and methodology section entails the questions :

1. Which parts of the dataset were you able to use, and how have you been able to use?
2. Which visualizations did you choose, why, what technologies (Python libraries, others) did you use for the visualizations ?

The indicative tasks of performing a map-based spatiotemporal analysis using choropleths has been completed. Heatmaps were intentionally omitted due to absence of relevant data for generating them. Previous inferences from Datathon 5 have been supported with new visualizations in addition to new inferences.

5. References

1. Sources from previous datathon have been utilized.
2. [List of countries by trade-to-GDP ratio](#)
3. [Commodity Dependence, Trade, and Growth](#)
4. [International trade in developing economies](#)
5. Book titled *Understanding Economic Development* by *NCERT* was additionally utilized.