**GitHUB - Launch project**

[**https://agatamalucha.github.io/UCDPA\_agatamalucha/index.html**](https://agatamalucha.github.io/UCDPA_agatamalucha/index.html)

**COVID PROJECT – HTML TEXT**

**Introduction**

Year 2020 was undoubtedly dominated by the pandemic caused by the new SARS-CoV-2 coronavirus. The pandemic has created many challenges for governments, enterprises and society globally. All countries in the world have been affected by the illness. Some of them have been hit harder than the others. As a healthcare worker who works in the hospital, I had great interest in all information about development of this disease, and I was especially concerned about situation in Europe.

**Problem description**

In every country in the world governments had to make difficult decision how to save public health from COVID illness. Were they successful? Which EU countries were hit the most with Covid cases and deaths? Did high numbers of hospital beds and healthcare workers helped in keeping low number of patient’s deaths? Is there any correlation between countries with low GDP and higher number of COVID cases? Did school closures and restriction of movement make any difference keeping the low number of cased? Were population with higher median age in worse situation through out the year of pandemic?

There are a lot of questions and I would like to find answers for some of them.

**Project Scope:**

The scope of “Covid cases in Europe” Project is to identify patterns, common elements among European Union countries with high and low number of Covid new cases and new deaths. This project will target health system elements, selected demographic, economical characteristics and internal countries policies.

In project scope:

* Project will target only current European Union countries.
* Project will target period of time between 2020-01-03 and 2021-05-03.
* Project will target only the following elements of demographics characteristics**:**
* GDP per capita (USD), median age.
* Project will target only the following elements of healthcare system characteristics**:**
  + hospital beds, nurses and midwives, physicians per 1000 population.
* Project will target only two specified internal restrictions:
  + School closures, internal movement restrictions for each day in the period, where "0" -no restrictions and "3"- highest restriction.

**Dataset sources:**

* **COVID cases and deaths dataset:**

Dataset have been provided by World Health Organization website in .csv format. Original dataset contains number of new cases, cumulative cases, and new deaths and cumulative deaths. Dataset has a minimum amount of missing data and contains worldwide daily data. In my project I use number of new cases and deaths and they will be recalculated to show cases and deaths per 10 000 population.

* **Economical dataset**:

Economical dataset of “GDP”, “population” have been sourced from the WorldBank website and imported as .csv files. Demographic data of “Median Age” has been scrapped from the website to show ability to find different HTML components, to use BeautifulSoup module and to import elements into .csv file for data analysis.

* **Healthcare dataset**:

Three major factors to be considered when analysing healthcare services in each European country are: available hospital beds, amount of nursing staff and number of physicians per 1000 population. Best website to acquire all necessary data is World Bank where the relevant data is being collected since 1960.

* **Internal policies dataset**:

Probably the most curious case and also the subject that create a lot of controversy impact of the lockdown policies and strategies on flattening COVID-19 cases curve.

After some research, I've found two relevant datasets. First one shows when schools

were closed in each European country, second one refers to movement restrictions.

Both of them can be found in "Our World in Data" dataset collection.

**Datasets analysing :**

**COVID cases and deaths dataset:**

Issues: Negative values --- Fill with “0” value

**Economical dataset** :

First task when pre-processing economical datasets was to remove non-EU countries, which was done with custom lambda function. Also, some country naming discrepancies were found (Czech Republic - Czechia). Some median age data was missing so dataframe had to be appended with relevant values. The whole cleaning process was finished with merging all datasets into one by country column. Visualizing dataset in a form of the bar chart give us some additional insights. You can easily observe that there are 16 countries where GDP range is between 10,000 and 30,000 USD/capita and 11 countries with GDP between 30,000

and 115,000 with much larger spread. When you also take for consideration the 75% centile, the amount (47614.01) is much lower that the max value and this shows that there are 2 high value outliers, Ireland and Luxembourg.

**Healthcare dataset:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Physicians\* | Nurses\* | hospital\_beds\* |
| mean | 3.586241 | 8.461981 | 4.998148 |
| std | 0.842889 | 3.247767 | 1.711239 |
| min | 1.9509 | 3.5534 | 2.21 |
| 25% | 3.0343 | 5.9515 | 3.335 |
| 50% | 3.4664 | 7.9665 | 4.69 |
| 75% | 4.01465 | 10.90885 | 6.59 |
| max | 5.4036 | 15.5735 | 8 |

\* per 1000 population

Top 5 counties with highest number of physicians, nurses and hospital beds per 1000 population:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TOP 3 Highest** | | | **TOP 3 Lowest** | | |
| **Physicians** | **Nurses** | **Hospital beds** | **Physicians** | **Nurses** | **Hospital beds** |
| 1. Greece | 1. Ireland | 1. Germany | 1. Cyprus | 1. Greece | 1. Sweden |
| 2. Austria | 2. Finland | 2. Bulgaria | 2. Romania | 2. Latvia | 2.Denmark |
| 3. Portugal | 3. Germany | 3. Austria | 3. Poland | 3. Bulgaria | 3. Ireland |
| 4. Lithuania | 4. Luxemburg | 4. Hungary | 4. Malta | 4. Cyprus | 4.Spain |
| 5. Germany | 5. Belgium | 5. Romania | 5. Luxemburg | 5. Spain | 5. Italy |

**Analysing Healthcare dataset:**

After removing non-EU countries and standardising their names, next step was to keep only the recent data column as datasets contain data from year 1960. Null value gaps were filed with most recent non-null data from previous years. Once all three datasets were merged into one by country column, I got curious if there is simple linear correlation between country's GDP and a size of a healthcare sector. Common sense would be to say, that wealthy country has more doctors than the poor one. So, I did some analysis in that subject.

**Plot Relation between GDP and Number of Physicians**

First step was to merge dataset with GDP column from economical dataset. Then I had to normalize the data to keep everything in range 0-1, where 1 is the top value and 0 is the lowest. This way it'll be easier for me to compare different value sizes. Once shown on the graph to the right, it's hard to say what top 10 countries have in common.

**Plot relation between GDP per capita and Nurses (Normalized)**

Much clearer situation is when it comes to nursing staff. Ireland takes the lead in Europe and generally, rich countries have more nurses than the poor ones. I think the main cause would be that the graduated nurses, often drastically underpaid are migrating for better job contracts.

**Plot relation between GDP per capita and Nurses (Normalized)**

Again, not that clear situation with hospital beds. But two conclusions can be made. First is that Post-soviet countries have more hospital beds due to the fact, that Soviet Union was preparing the invasion on the West and those countries supposed to be the battlefield. Secondly interesting observation is Sweden - I think that has a very low number of hospital beds forced government to seek herd immunity instead of taking a path of the other European countries.

**Dataset Policies Analysis**

Here is the most controversial part of my analysis. First task was to transform the dataset the way, that we could see the number of days per restriction level per country. In the future analysis I'll try to observe if there is any correlation between GDP and number of days in lockdown. Also, I'll try to measure how effective was each restriction and if its quick application prevented case spikes.

**Datasets Insights**

In this part of my project, I would like find out how all European union counties have managed Covid in analysed period of time by comparing total number of Covid cases and deaths per 10,000 population with countries own characteristics.

**Total Covid cases vs GDP.**

Fist of all I wanted to check if there is correlation between number of cases in each country and their GDP per capita. In this investigation I haven’t noticed any obvious visual correlations. The only thing that is noticeable is 4 out of 5 the richest counties (highest GDP per capita, please see Economical Dataset analysis) Ireland, Denmark, Finland has the total lowest number of COVID cases.

**Total Covid deaths vs Median Age.**

In this pair of features you can notice not definite but noticeable relation. Number of Covid deaths is slightly falling as falling the median age of population in the country. Countries with higher median age experienced higher number of Covid cases in total.

**Total Covid deaths vs Number of days in the highest school closure restriction.**

On this chart I can see counties with number of days higher than average closed schools which have experienced high number of deaths (Czechia, Slovenia, Slovakia). On the other side of chart there are countries like Latvia, Germany, Ireland, Greece, Malta with also high number of school closed days but less than average deaths. I will look into those counties in final chart, as those group of countries might have more things in common.

**Number of Nursing staff vs Severity in restrictions with normalized data.**

On this chart there is visible relation between number of nursing staff and number of days in school/ internal movement lockdown. You can see that countries with high level of staffing have experienced lower level of restriction as number of nurses is falling countries implemented longer and more server lockdown. There is one more intresting fact that I have noticed that when I taking into account GDP per capita , wealthy countries Ireland, Germany , Finland , Luxemburg have high number of staff.

**Total Covid deaths vs Number of days in the highest internal movement and school closure restriction with normalized data.**

In this chart I compared new number of deaths with number of days in both internal movement restriction and school closures. Again, the worst situation can be observed in eastern European countries even though they had longer /higher level of restrictions. On the other hand we have Finland , Denmark, Malta, Estonia. Even though they had no restrictions they had low number of deaths. Ireland , German and Greece kept their restriction for longer and that helped them to keep number of deaths lower than average in EU.

**Daily Covid cases vs. school closures and movement restrictions.**

I am continuing analysing impact of policies on course of Covid cases through out the whole period of pandemic. Covid was a new disease , so as precaution all European countries closed schools and restricted internal movements , even though there was a very low number of new Covid cases. Europe have started extreme lockdown. This can be observe between March 2020 and May 2020 . As we were learning more about Covid and reporting low number of cases,all EU started ceasing on restrictions. You can see that particularly between July-Sept 2020. The next turning point happened in October/November 2020 , Covid cases started dramatically increase. Countries decided to introduce lockdowns again, however in less strict conditions. Prolonged period of lockdown at the start of pandemic , probably has put counties in economic trouble and also politicians started to encounter society resistance on severe lockdown. From January 2021 up to the current date the number of cases remained quite high with higher level of restrictions. Only in April /May 2021 countries have dropped restrictions as they strongly think that implementation of vaccination against Covid will suppress the disease.