A picture containing cake, table, white

Description automatically generated

**Project #1  
COVID-19: Data Research**

Analysis Document

**Equipo F:**

Cecilia Acosta | Luis Bonilla | Felipe Murillo

**May 16, 2020**

TEC Monterrey Data Analytics Bootcamp

## **Motivation**

Currently the world is facing a pandemic on an unprecedented scale. At the end of 2019 and grossly picking up momentum in 2020, the world is now a different place after COVID-19 and the SARS-CoV-2 virus that causes it. As of May 2020, COVID-19 has affected over 4M people worldwide and has claimed over 300K lives.

The death toll does not yet compare to other well-known pandemics in our history, such as the Black Death (Bubonic Plague) of 1347 – 1351 A.D. which caused ~200M deaths, or the Spanish Flu of the early 1900s which claimed between 40-50M deaths. None of us would have imagined that in our lifetimes, we would experience an event like this one.

There’s a lot of information swirling around this topic, and our team was motivated to carry out a deeper analysis to draw our own conclusions about COVID-19 cases, mainly from a correlation perspective.

## **Collected Data**

To explore how COVID-19 has affected the world from a geographic and income perspective, we decided to focus our correlation effort around the following monetary and health measures:

For that it was necessary retrieve information from trusted sources such as the World Bank, World Health Organization (WHO), and the John Hopkins Institute:

* **COVID-19 mortality rate**
  + Calculated proportion of recorded deaths to total cases, per country
  + Source: World Health Organization (<https://www.who.int/es>)
* **Hospital Beds (per 10,000 individuals)**
  + Normalized measure of how well-equipped a country is to treat the virus. There is a known issue now in hospitals who are running out of capacity to treat patients with COVID-19.
  + Source: World Health Organization (API)

([http://apps.who.int/gho/athena/api/GHO/ WHS6\_102.json?profile=simple](http://apps.who.int/gho/athena/api/GHO/%20WHS6_102.json?profile=simple))

* **GNI per capita**
  + Gross national income (GNI) per individual, a measure of standard of living
  + Source: World Bank (API)

([http://api.worldbank.org/v2/country/*<country>/*indicator/ NY.GNP.PCAP.CD?format=json](http://api.worldbank.org/v2/country/%3ccountry%3e/indicator/%20NY.GNP.PCAP.CD?format=json))

* **Health expenditures (in USD, % of GDP)**
  + How much a country is spending on healthcare
  + Source: World Bank
  + Out-of-pocket expenditure (% of current health expenditure). (n.d.); Retrieved May 15, 2020: (<https://data.worldbank.org/indicator/SH.XPD.OOPC.CH.ZS>)
  + Current health expenditure (% of GDP). (n.d.); Retrieved May 15, 2020:

(<https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS>)

Once we call the datasets of the different sources mentioned above we decided to focus on relationships between the Mortality Rate and Hospital Beds Available and Economic Indicators (GNI per Capita, Health Expenditure per Capita and as % of GDP)

## **Analysis**

Is this pandemic impacting nations indiscriminately or are certain regions affected more than others? Should lower income countries see more COVID-19 cases and deaths? Are there patterns that can be identified?

Mapping out COVID mortality rates around the world shows hotspots in Europe and regions in the Caribbean & Central America. Data shows mortality rates are higher in countries with higher GNI per capita. This could be explained by 2 factors:

1. Increased international travel of citizens of these countries may come with increased risk of virus exposure (in general terms, higher GNI nations have higher purchasing power as compared with other regions).
2. Countries with more resources can achieve more testing, as shown for example by Germany.

## Are there specific correlations between mortality rates and factors such as health expenditure, number of hospital beds and GNI?

The data supports the fact that there are no correlations between COVID-19 mortality rates and our selected variables, even at the geographic and income group levels. Pearson correlation coefficients associating mortality rate to health expenditure, GNI per capita and hospital beds do not indicate a significant association. The effects of COVID-19 in a given region/income group can NOT be predicted by the indicators used in our study.

## What is the correlation of Hospital Beds vs. COVID-19 cases reported?

**By Region:** The results show that regions with higher income, have better hospital bed availability. For instance, Europe (a known High Income region) has 43% more hospital beds than the Americas. One would expect worse COVID mortality in nations with less beds; but that’s not the case. Nations with good Hospital Services are experiencing higher stresses during this COVID-19 situation, which should be alarming for regions with fewer resources not yet peaking in their COVID cases (future impact could be catastrophic!).

**By Income**: Data clearly demonstrates that there are more hospital beds available in high income regions, but there are also 40% more cases. Reasons? There 2 main hypotheses are:

1. Low income regions do not have the resources to perform COVID-19 tests at a wider-scale, so there might be several unrecorded cases that deviate the statistics.
2. High income countries have more possibilities to carry international travel, directly related to the purchasing power of their citizens.

## What is the correlation of Hospital Beds vs GNI?

**By Region:** The results show that with exception of Europe and Asia, the GNI does not have a clear relation with available hospital beds. That does not mean they are not spending enough money on health, but the split between concepts may be different (drugs, infrastructure, payroll of doctors and nurses, etc.)

**By Income**: This analysis brings a cold data due the delta of GNI between Upper Middle to High Income groups. The increase is up to 69% which shows that the Upper Middle bucket is closer to the Low Income than to the High income. However, the expenditure in Hospital Beds has an almost flat increase from Low Income to High Income.

## Are nations with high health expenditures and more hospital beds having lower mortality rates?

Countries with the highest health expenditures do not necessarily have the most number of beds, which takes us to the conclusion that the gross spending is carried out in other concepts which could be research and development, payroll of doctors and nurses, drugs, infrastructure, etc. This translates in different mortality rates among countries of different geographies and income levels. For instance, France, Sweden, Belgium, Netherlands, UK, Spain, and Italy are amongst the highest GNI and health expenditures as % of GDP, but also show high mortality rates above the mean of 5% (> 10%).

## **Conclusions**

The biggest inference based on the analyzed data is that there is no significant correlation between mortality rates and health expenditures or income levels. All regions have the same probability of infection and deaths. Initially, one of the hypotheses was that high income countries will spend more on healthcare will have the lowest mortality rates but that proved not to be the case. The difference with low income countries is high so this leads to the conclusion that these countries have less resources to spend on acquiring COVID-19 tests and therefore they do not report high numbers of cases or deaths, which makes them less comparable. Even within Europe, one of the most impacted regions, Germany carried extensive tests and reported more cases than Spain or Italy, 2 of the countries with the highest mortality rates.

Team F