INDEX

Project Description

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (China-WHO Joint Mission, 2020). The disease was first identified in 2019 in Wuhan, the capital of Hubei China, and has since spread globally, resulting in the 2019–20 coronavirus pandemic.

In this project, we perform a comparative analysis of COVID-19 expansion across **Spain**.
The visual representation and analysis of COVID-19 patients in Spain is shown.
Rise and drop of number of patients in different areas of Spain is highlighted.
Maximum affected areas are marked.
The impact of lockdown and COVID-19 on People's mobility is shown.

Problem Statement

City wise comparison, in which city there were maximum cases and which city had minimum. In addition, in which cities the number of cases increased highly in short span of time?

In how many days the number of cases increased drastically or higher number of cases were reported?

How many deaths were reported in what number of days? When did the recovery rate increase?

Research Papers

We have referred to the following research papers:

http://www.amegroups.com/
http://biotech-spain.com/en/articles/evidence-covid-19/
https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(20)30060-8/fulltext

Data Source

Official Spain covid-19 website - http://covid19.isciii.es
John Hopkins University and Spain Ministry of Health

Data Description

Name	Description
CCAA	Autonomous communities in Spain.
Fecha (Date)	Date, starting from 20th February.
Casos (Cases)	Number of cases on the particular day in the specific community.
PCR+(tests)	The PCR (polymerase chain reaction) test results in patients recovered from Covid-19.
TestAc+	Antibody confirmed test, which is the number of patients who need intensive care unit.
Hospitalizados (Hospitalised)	Number of patients who have been hospitalised.
UCI (ICU)	Number of patients who need intensive care
Fallecidos (Deceased)	Number of patients died.
Recuperados(Recovered)	Number of patients who have recovered completely

Technologies Used

- Python : Detailed graphical visualizations for showing different aspects in Spain such as, number of cases/deaths/hospitalized with respect to date.
- Tableau : Designing and layout of Dashboard.
- Html, CSS : Basics for creating the website link.

Data Pre-processing

Firstly, we install GeoPandas as it helps to make working with geospatial data in python much easier. We import libraries like numpy, pandas, Matplotlib, etc. for further use.
To load the data, we use the traditional 'read_csv' method mentioning the URL of our dataset as the file path.
The CSV file has last eight lines in Spanish, which are not useful so we get rid of it by deleting it.

Then, we replace the NULL values with zero.
The data set column names are in Spanish. Therefore, we change it into English.
Then, we change the abbreviations of the states (communities) to their full forms.
We create new columns where we have the count of daily infected, deaths and cases.

Visual Exploration

1. Types of visualisation used – What kind of Information it depicts.

    1. Daily infection, deaths in top 6 states of Spain with time.
    The top 6 infected states are (first) Madrid, Cataluña, Valencia, Castilla La Mancha, Castilla y Leon, País Vasco (sixth).

    2. Mortality rate in different states of Spain with date.
    This statistic presents the COVID-19 mortality rate per 100,000 inhabitants in the Spanish autonomous communities, as of May 13, 2020. Madrid was the most affected by the coronavirus, both in terms of mortality rate, which stood at 132 deaths per 100,000 people; and in number of cases, which was 65 thousand as of the same date.
    (Mortality rate is a measure of the number of deaths in a particular population, scaled to the size of that population, per unit of time. Mortality rate is typically expressed in units of deaths per 1,000 individuals per year)

    3. Infected, ICU, Hospitalized, Recovered, Deaths over time.
    The virus affected different states differently since February until now.
    Madrid was affected the most. Although, all curves are exponentially increasing.

    4. Infected, ICU, Hospitalized, Deaths with respect to ages.
    The virus mostly affected older people. The death´s curve peak is around 80-89yo.
    The total infected line as the 'peak' goes from 40 to 80 years, which means the disease is present in most of the population.

    5. Total infected with respect to gender.
    The virus is more dangerous in men than woman although the total infected people is higher in woman until 60yo. However, the age ranges in which women are more infected than men are the less dangerous ages. There are more men than woman infected in the dangerous ages (+60yo).

2. Technologies used for visualization.

    - Python
    - Tableau

Model Building

1. Objective.

   To analysis and predict the recovery rate in Spain and the different autonomous communities.

2. Algorithm used.

   The SIR Model is used in epidemiology to compute the amount of Susceptible, Infected, and Recovered people in a population. It is also used to explain the change in the number of people needing medical attention during an epidemic.

   The SIR model divides the (fixed) population of N individuals into three "compartments" which may vary as a function of time, t:

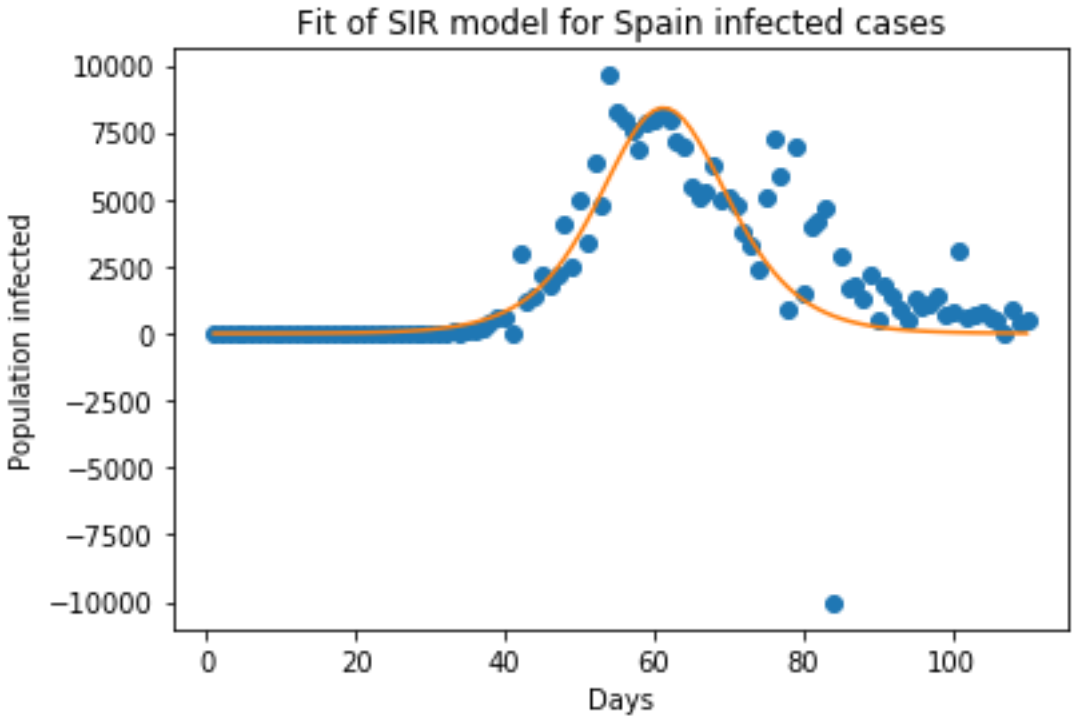
- S(t) are those susceptible but not yet infected with the disease;
- I(t) is the number of infectious individuals;
- R(t) are those individuals who have recovered from the disease and now have immunity to it.

The SIR model describes the change in the population of each of these compartments in terms of two parameters, β and γ. β describes the effective *contact rate* of the disease: an infected individual comes into contact with βN other individuals per unit time (of which the fraction that are susceptible to contracting the disease is S/N). γ is the mean recovery rate: that is, 1/γ is the mean period of time during which an infected individual can pass it on.

$$\frac{dS}{dt} = -\beta \cdot I \cdot \frac{S}{N}$$
$$\frac{dI}{dt} = \beta \cdot I \cdot \frac{S}{N} - \gamma \cdot I$$
$$\frac{dR}{dt} = \gamma \cdot I$$

Spain has the same number of cumulative infected cases than Italy, near 120.000. However, Spain's total population is lower (around 42 millions) and hence the percentage of population that has been infected rises up to 3%.
SIR is a simple model that considers a population that belongs to one of the following states:

Susceptible (S). The individual has not contracted the disease, but she can be infected due to transmission from infected people Infected (I). This person has contracted the disease Recovered/Deceased (R). The disease may lead to one of two destinies: either the person survives, hence developing immunity to the disease, or the person is deceased.



Optimal parameters: beta = 9.15267663109941 and gamma = 8.979627255691696

Models tailored specifically for epidemic spreading (i.e. SIR and its versions) are designed to give a certain idea, in order to understand the spread of the disease. On the other hand, the simple machine learning approaches we used, aim to predict the short-term evolution of the infection in the current scenario. They might eventually help to find some features or parameters that are particularly important for the model's fitting, but by no means they should be confused with scientific epidemic models. We are trying to refine the model but due to lack of complete information, we are unable to build the model perfectly.

Analysis

Spain had a higher mortality rate among those aged over 90 years old. 22% of these that contracted the COVID-19 died from the complications caused by this virus. This disease mostly affected those aged 70-79 years with about 23% of coronavirus hospitalizations found in that age range.

The number of people affected by this coronavirus in Spain was about over 235 thousand. Madrid, the Spanish capital is the region with the highest number of cases, and had a transmission rate of 1,074 cases every 100 thousand people.

Total Population: 46,753,345
Total coronavirus cases: 286,718
Total coronavirus deaths: 27,127
Total recovered cases: 196,958
That is, Coronavirus affect approximately 0.6% of the population and around 9% of that have died.
(As per 1 June)

Observations

- Firstly, we observed that the first COVID-19 case in Spain was on 20 February 2020 in Balearic and Canary Islands.
- Spain was one of the hardest-hit countries in the pandemic, the coronavirus spread quickly and widely without being detected, especially among the elderly.
- In addition, we observe that the curve is normally distributed that is, number of cases decrease after an interval of time.
- Madrid and Catalonia are the two hardest-hit communities.
- Also, we observe that (18th April)
- Spain has the lowest daily death count compared to the other countries.
- We also observe what the impact of lockdown and the coronavirus has on people's mobility. Pharmacy and Grocery is used more whereas people's mobility at
- According to us, the spread of the virus is due to the following reasons :
    - The country's sociable lifestyle and close ties between young and older family members.
    - The houses are located near each other.
    - Spain has a large elderly population.
    - Freedom of movement during the outbreak. As the Spanish authorities implemented a strict lock-down on March 14 when the number of cases was 6391.

Future Enhancement

This project has a vast scope in the Future. It can be updated in near future as and when requirement for the same arises, as it is very flexible in terms of expansion.
The data would be updated time to time, which would help for better analysis.
The data is ongoing so there were a few difficulties to make graphical visualizations.
As few of the parameters were incorrect and incomplete, this can be corrected with time.
A better and trusted data source could be used.

Different part of the world was hit differently by the corona virus, which further affected the different sectors such as manufacturing, agriculture, IT, education etc. Hence, to avoid such a pandemic in the near future, effective measures will be taken.

Bibliography

- https://github.com/CSSEGISandData/COVID-19
- https://www.worldometers.info/coronavirus/country/spain/
- https://cnecovid.isciii.es/covid19/
- https://www.kaggle.com/danigarci1/covid19-in-spain