**BLUEEDTECH**

**ANA CRISTINA CHAVES**

**ANTONIO DUARTE JUNIOR**

**THIAGO CHARLES**

## **BOOTCAMP - PESQUISA DE IMPACTO DO CORONAVIRUS 2020**

**BRASIL**

**NOVEMBRO 2022**

**INTRODUÇÃO**

### 

### Uma pandemia é uma epidemia de doença infecciosa, que se espalha entre a população localizada numa grande região geográfica como, por exemplo, todo o planeta terra.

Em 31 de dezembro de 2019, a Organização Mundial da Saúde (OMS) foi alertada sobre vários casos de pneumonia na cidade de Wuhan, província de Hubei, na República Popular da China. Tratava-se de uma nova cepa (tipo) de coronavírus que não havia sido identificada antes em seres humanos.

Uma semana depois, em 7 de janeiro de 2020, as autoridades chinesas confirmaram que haviam identificado um novo tipo de coronavírus. Os coronavírus estão por toda parte. Eles são a segunda principal causa de resfriado comum (após rinovírus) e, até as últimas décadas, raramente causavam doenças mais graves em humanos do que o resfriado comum.

Com a pandemia, o número de estudos aumentou drasticamente e junto o volume de dados produzidos e distribuídos por todos os países com objetivo de ajudar a combater o vírus também sofreu um aumento considerável.

Para melhor controle da doença, foi gerado muitos dados estatísticos, muita informação, para combate, para prevenção, enfim, os dados foram utilizados da melhor forma em meio ao caos, e agora que a pandemia deu uma trégua, iremos utilizar as informações contidas nesses dados reunidos, referente a cinco países Espanha, equador, Chile, México e Argentina em um intervalo de Janeiro a Julho de 2020.

**CONJUNTO DE DADOS**

**Conjunto de dados COVID-19** *, Kaggle* Este conjunto possui dados do número de casos confirmados, mortes e recuperados por dia em vários países entre Janeiro de 2020 a Julho de 2020.

* country\_wise\_latest.csv
* Covid\_19\_clean\_complete.csv
* day\_wise.csv
* Full\_grouped.csv
* Usa\_county\_wise.csv
* worldometer\_data.csv

**API do Twitter** A rede social Twitter.

**WebScraping** Utilização da técnica de garantia de dados de sites, como Wikipédia.

**DICIONARIO DE DADOS**

|  |  |
| --- | --- |
| **worldometer\_data.csv** | |
| Country/Region | País/Região |
| Continent | Continente |
| Population | População |
| TotalCases | Total de casos |
| NewCases | Novos casos |
| TotalDeaths | Total de mortes |
| NewDeaths | Novas mortes |
| TotalRecovered | Total de recuperados |
| NewRecovered | Novos casos recuperados |
| Active Cases | Casos ativos |
| Serius, Critical | Casos sérios/ críticos |
| Tot Cases/1m pop | Total de casos por 1m habitantes |
| Death/ 1m pop | Morte por 1m de habitantes |
| TotalTests | Total de testes |
| Tests/1M pop | Testes por 1m de habitantes |
| Who Region | Região da OMS |
|  |  |

|  |  |
| --- | --- |
| **covid *19* clean\_complete.csv - Dia a dia do país não. de casos (não possui dados em nível de condado/estado/província)** | |
| Province/States | Províncias e estados |
| Country/Region | País/Região |
| Lat | Latitude |
| Long | Longitude |
| Date | data |
| Confirmed | Confirmados |
| Deaths | Mortes |
| Recovered | Recuperados |
| Actives | Ativos |
| Who Region | Região da OMS |

|  |  |
| --- | --- |
| **full\_grouped.csv - Dia a dia do país não. de casos (tem dados em nível de condado/estado/província)** | |
| Date | Data |
| Country/Region | País/Região |
| Confirmed | Casos confirmados |
| Deaths | mortes |
| Recovered | recuperados |
| Active | Casos ativos |
| New Cases | Novos casos em 24h |
| New deaths | Novas mortes |
| new recovered | Novos casos recuperados 24h |
| who region | Região da OMS |

|  |  |
| --- | --- |
| **country *wise* later.csv - Nº do nível do país mais recente. de casos** | |
| Country/Region | País/Região |
| Confirmed | Casos confirmados |
| Deaths | mortes |
| Recovered | recuperados |
| Active | Ativos |
| New cases | Novos casos em 24h |
| New deaths | Novas mortes |
| New recovered | Novos casos recuperados 24h |
| Deaths / 100 Cases | Mortes/ 100 casos |
| Recovered / 100 Cases | Recuperados/ 100 casos |
| Deaths / 100 Recovered | Mortes / 100 Casos |
| Confirmed last week | Confirmados na semana passada] |
| 1 week change | Mudança de 1 semana |
| 1 week % increase | semana % de aumento |
| WHO Region | Região da OMS |

|  |  |
| --- | --- |
| **day\_wise.csv - Dia sábio não. de casos (não possui dados em nível de país)** | |
| Não iremos utilizar por não ser específico por país como foi solicitado | |
|  |  |
| **usa *county* wise.csv - Dia a dia no nível do condado. de casos** | |
| Informações somente dos Estados Unidos | |

## Análises de dados

Covid clean métricas mundiais – Thiago

Country e full - ana

## Metodologia

Faremos uma análise das séries dos dados.

Modelos a ser testado ARIMA para serie temporais.

Long Short Term Memory (LSTM).

Métrica MAE, RMSE, R2, por ser um problema de regressão.

## Dashboard

Os dados gold serão explorador no PowerBI, informações com base em taxa de incidência por 100mil/habitantes e podemos verificar pelo PIB no País.

Este rastreador fornece o número cumulativo de casos e mortes confirmados de COVID-19, bem como a taxa diária de casos e mortes de COVID-19 por país, renda, região e globalmente. Este rastreador será atualizado regularmente, à medida que novos dados forem divulgados. Consulte a guia Métodos para obter informações mais detalhadas sobre fontes de dados e notas.

## Referência

Barreto, Evandro Fernandes Ciência de Dados aplicada à pandemia do coronavirus no Brasil, uma análise socioeconômica. Universidade Estadual Paulista (Unesp), 2021. Disponível em: <<http://hdl.handle.net/11449/213821>>.