



A eficácia da Vacina na Redução de Mortes por Covid 19 no estado de SP

Ivy de Moraes Tentoni

Pós Graduação Latu Sensu em Ciência de Dados e Big Data
Pontifícia Universidade Católica de Minas Gerais
17/04/2022

Contextualização

Em 2019 o mundo todo viveu algo que não imaginávamos viver em pleno século 21, fomos surpreendidos por uma Pandemia que chegava em alta velocidade e mudando tudo.

A COVID-19 surgiu no Brasil com seu primeiro caso registrado em 26 de Fevereiro de 2020 , é uma doença infecciosa causada pelo coronavírus SARS-CoV-2, a Organização Mundial da Saúde (OMS) tomou conhecimento sobre o novo Vírus em 31 de Dezembro de 2019

Logo que surgiu a primeira “Onda” de Covid no Brasil vivenciamos um cenário caótico no mundo todo , muitas dúvidas sobre a doença, informações falsas, hospitais lotados, escolas fechadas, shoppings fechados, comércios fechados, restaurantes fechados tudo “parou”.

Vimos ruas vazias, desemprego aumentando, vimos as pessoas que podiam trabalhar em casa se adaptando a essa nova realidade, empresas se adaptando para continuar produzindo de forma remota e para aqueles que atuavam em serviços essenciais ou atividades que não tinham como ser realizada de forma remota se arriscando para garantir o sustento de suas famílias.

Certamente um período de incertezas, medos, perdas, muitas mudanças um período que não será esquecido.

Porém em meio ao caos surgiu uma nova esperança a Vacina e tudo mudou novamente!

Problema Proposto

O Objetivo deste trabalho é comprovar com análise de dados de que após o Início da campanha de Vacinação houve a redução do caso de Óbitos causados por Covid-19.

Gostaria de colaborar com a população divulgando informação para estimular a população a utilizar dos dados para auxiliar no processo de decisão.

Problema Proposto – 5Ws

(Why?) Por que esse problema é importante?

- Este problema é importante para conscientizar o cidadão da importância de tomar a Vacina contra a Covid.
- A vacina visa reduzir o risco de pegar COVID-19 e evitar quadros graves e potencialmente fatais.

(Who?) De quem são os dados analisados? De um governo? Um ministério ou secretaria? Dados de clientes?

- Dados do Governo Brasileiro, disponibilizado gratuitamente e de forma aberta no dataset (<https://brasil.io/dataset/covid19/caso/>)
- Em específico será analisado os dados do Estado de SP.

(When?): Qual o período está sendo analisado? A última semana? Os últimos 6 meses? O ano passado?

(What?): Quais os objetivos com essa análise? O que iremos analisar?

- O Objetivo desta análise é comprovar com dados de que após o Início da Vacinação houve a redução do caso de Óbitos por Covid-19, comprovando portanto a efetividade da vacina.

(Where?): Trata dos aspectos geográficos e logísticos de sua análise.

- A base tem dados de todo Brasil , porém para análise em questão será utilizado os Dados apenas do estado de SP

- Serão analisados dados de 2020 até 2022

Coleta de Dados

O Dataset utilizado para o desenvolvimento desta pesquisa foi obtido através da página Brasil.io , página do qual tem por missão tornar acessível os dados brasileiros de interesse público e que tem como valor principal a transparência e colaboração.

Efetuei o download em meu computador do dataset completo do arquivo “caso_full”, o formato da extração é em um arquivo CSV e após extração os dados serão tratados e segregados ao longo do desenvolvimento do trabalho.

O dataset original possui 18 colunas e 38.536.48 linhas

```
1 import pandas as pd
2
3
4 dados = pd.read_csv(r'C:\Users\ivymo\OneDrive\Área de Trabalho\TCC\caso_full.csv')
5
6 print(dados)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

Copyright (C) Microsoft Corporation. Todos os direitos reservados.

Instale o PowerShell mais recente para obter novos recursos e aprimoramentos! <https://aka.ms/PSWindows>

PS C:\Users\ivymo> & C:/Users/ivymo/AppData/Local/Programs/Python/Python318/python.exe "c:/Users/ivymo/AppData/Local/Programs/Python/Python318/Scripts/pd-read_csv.py" "C:\Users\ivymo\OneDrive\Área de Trabalho\TCC\caso_full.csv"

	city	city_ibge_code	date	epidemiological_week	estimated_population	...
0	Rio Branco	1200401.0	2020-03-17	202012	413418.0	...
1	NaN	12.0	2020-03-17	202012	894470.0	...
2	Rio Branco	1200401.0	2020-03-18	202012	413418.0	...
3	NaN	12.0	2020-03-18	202012	894470.0	...
4	Rio Branco	1200401.0	2020-03-19	202012	413418.0	...
...
3853643	Tupirama	1721257.0	2022-03-27	202213	1922.0	...
3853644	Tupiramins	1721307.0	2022-03-27	202213	2729.0	...
3853645	Wanderlândia	1722081.0	2022-03-27	202213	11734.0	...
3853646	Xambioá	1722107.0	2022-03-27	202213	11520.0	...
3853647	NaN	17.0	2022-03-27	202213	1590248.0	...

[3853648 rows x 18 columns]

PS C:\Users\ivymo>

Processamento/Tratamento de Dados

Como a base de dados é muito grande efetuei primeiramente o download do arquivo em minha máquina local e posteriormente o import. do arquivo CSV no Python e utilizando a Biblioteca Pandas analisei a quantidade de linhas e colunas por estado.

A base de dados não necessitou de uma tratativa para dados duplicados e/ou ausentes o dataset já foi devidamente tratado pelos seus autores.

Fiz uma alteração nas colunas modificando o nome das colunas New Deaths para Óbitos Confirmados e New Confirmed para Casos de Covid Confirmados.

Observando o resultado do total de linhas decidi focar a análise em um estado específico o segundo estado com maior volume de linhas (ou seja, dados registrados) para uma análise mais direcionada, o estado escolhido para minha análise foi SP.

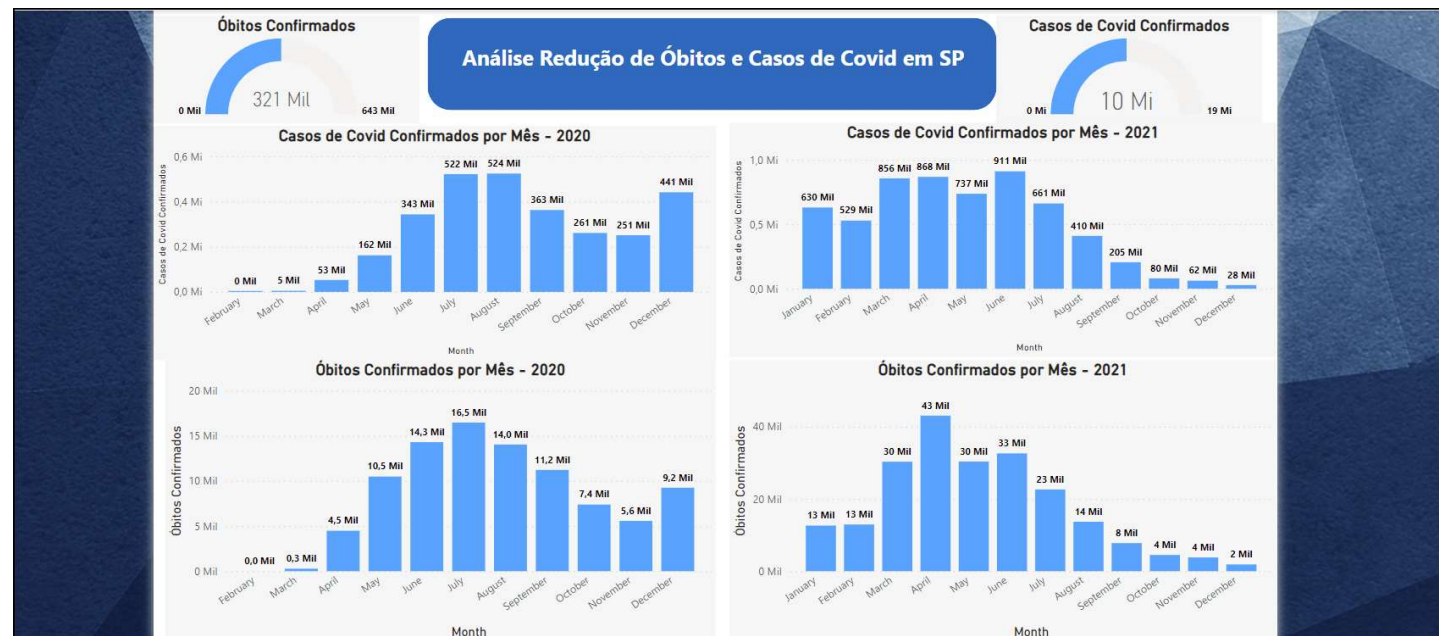
Em análise do dataset Casos_full.csv o estado de SP possui a seguinte quantidade de linhas:

Total de Linhas:4.489.23

Estado	Total de Linhas
Minas Gerais (MG)	574802
São Paulo (SP)	448923
Rio Grande do Sul (RS)	335719
Bahia (BA)	287758
Paraná (PR)	272487
Santa Catarina (SC)	204645
Goiás (GO)	167029
Paraíba (PB)	154071
Piauí (PI)	152463
Maranhão (MA)	151588
Ceará (CE)	131734
Pernambuco (PE)	131561
Rio Grande do Norte (RN)	116993
Pará (PA)	102768
Mato Grosso (MT)	96994
Tocantins (TO)	95770
Alagoas (AL)	72120
Rio de Janeiro (RJ)	67452
Espírito Santo (ES)	56954
Mato Grosso do Sul (MS)	54377
Sergipe (SE)	52923
Amazonas (AM)	44620
Rondônia (RO)	37388
Acre (AC)	16150
Roraima (RR)	12064
Amapá (AP)	12061
Distrito Federal (DF)	2234

Análise/Exploração dos Dados

Entendendo o volume de dados, definindo um foco para a análise decidi utilizar o PowerBi para a criação de um dashboard e assim analisar os dados de uma forma gráfica o que possibilitaria ilustrar de uma forma mais prática o propósito do meu trabalho que é provar com dados a redução do número de Mortes por Covid após a campanha massiva de Vacinação



Análise/Exploração dos Dados



Gráfico : Casos de Covid Confirmados

Contém o Total de Casos Confirmados por Covid em SP
Analisando esta informação já é possível observar que de um total de 19 milhões, SP representa 52,63% do total 10 milhões de casos registrados.
Este campo representa a Soma do campo "Casos Covid Confirmado"



Gráfico : Óbitos Confirmados

Contém o Total de Óbitos Confirmados por Covid em SP
Analisando esta informação já é possível observar que de um total de 643mil casos registrados 321mil foram registrados em SP cerca de 49,92%.



Gráfico: Casos de Covid confirmados por Mês

Contém uma visão de quantidade de casos de Covid registrados por Mês no ano de 2020

Este gráfico representa a Soma do campo "Casos Covid Confirmado" mais um adicional do filtro de data pelo ano de 2020 mais a soma dos casos por Mês para facilitar a visão e análise.

Este gráfico já permite visualizar claramente como a Pandemia se agravou de uma forma muito rápida no estado de SP, partimos de 5 mil casos registrados em março para 522 mil em julho em apenas 5 meses.

Análise/Exploração dos Dados



Gráfico : Casos de Covid por Mês 2021

Este gráfico representa a Soma do campo “Casos Covid Confirmado” mais um adicional do filtro de data pelo ano de 2021 mais a soma dos casos por Mês para facilitar a visão e análise

Este é um dos gráficos mais importantes do dashboard, claramente perceptível a redução de casos de covid após julho de 2021 quando a vacina chegou para a maior parte da população adulta em SP.



Gráfico : Óbitos confirmados por Mês – 2020

Este campo representa a Soma do campo “Óbitos Confirmados” mais um adicional do filtro de data pelo ano de 2020 mais a soma dos casos por Mês para facilitar a visão e análise.

Este gráfico já demonstra o pico que tivemos na Pandemia nos meses de junho e julho



Gráfico : Óbitos confirmados por Mês 2021

Este gráfico representa a Soma do campo “Óbitos Confirmados” mais um adicional do filtro de data pelo ano de 2021 mais a soma dos casos por Mês para facilitar a visão e análise.

Este sim é o gráfico mais importante de todo o dashboard ele é o resultado da análise que estou propondo para este trabalho, é possível ver claramente a redução do número de Óbitos confirmados a partir de Outubro, cenário totalmente o contrário do que vimos no ano de 2020.

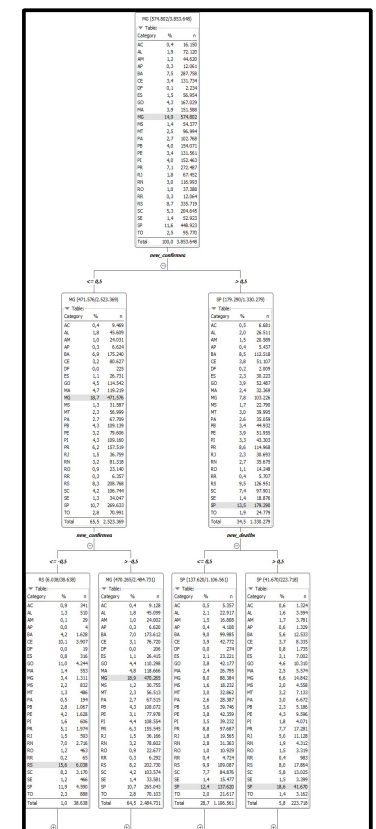
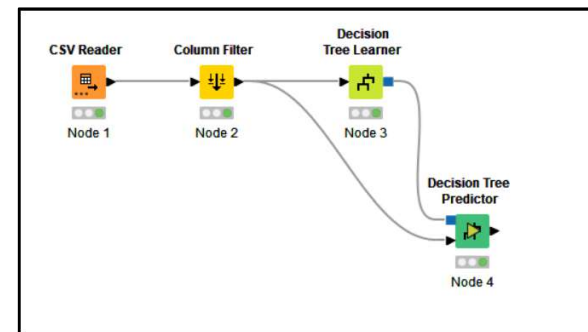
Redução de 40% dos casos comparando Outubro de 2020 com Outubro de 2021
Redução de 28,57% dos casos comparando Novembro de 2020 com Novembro de 2021
Redução de 78.26% dos casos comparando Dezembro de 2020 com Dezembro de 2021

Criação de Modelos de Machine Learning

Para criação do Modelo de Machine Learning deste trabalho utilizei o Knime.

O algoritmo de Machine Learning escolhido foi o da Árvore de Decisão, utilizei a árvore de decisão para avaliar a tomada de decisão pelo estado escolhido pela análise e também do estado com o maior casos de óbito registrado.

A árvore de decisão é um algoritmo de aprendizado e máquina supervisionado que é utilizado para classificações e regressão muito similar a um fluxograma com etapas fáceis de visualizar e entender além de ser recomendada para uma compreensão melhor de como funciona o aprendizado de máquina.



Criação de Modelos de Machine Learning

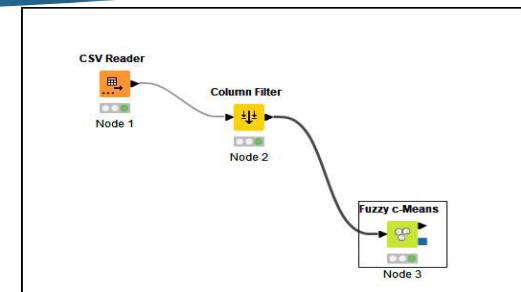
Efetuei também a criação de um novo fluxo utilizando o algoritmo Fuzzy c-Means, este algoritmo utiliza uma forma de agrupamento de elementos demonstrando assim qual elemento pode pertencer a mais de um grupo.
Antes de gerar o fluxo deste algoritmo efetuei a geração de uma nova base de dados utilizando o mesmo dataset “casos_full” efetuei um agrupamento com as seguintes informações:

Ano
Mês
Soma Casos de Óbitos Confirmados
Soma Casos de Covid Confirmados
Filtro por estado de SP

A partir desta informação efetuei o export. dos dados para analisar no Knime, gerando assim um novo arquivo covidanalyse.csv

O algoritmo Fuzzy C Means é uma técnica bem conhecida de aprendizado não supervisionado que releva a estrutura subjacente dos dados , o agrupamento do fuzzy permite que cada ponto de dados pertença a vários clusters, com um grau de pertinência a cada um

Utilizando este algoritmo é possível ter uma visão clara de que os Meses do qual a campanha de vacinação já estava mais avançada no país reduzindo o registro de Óbitos por covid, estes casos foram categorizados no Cluster 2, com exceção dos casos de Março e Fevereiro de 2020 pois estes meses foram o inicio da pandemia no país



Row ID	Year	Month	Óbi...	Estado	cluster_0	cluster_1	cluster_2	Winner Cluster
Row0	2020	February	0	SP				cluster_2
Row1	2020	March	272	SP				cluster_2
Row22	2021	December	1858	SP				cluster_2
Row25	2022	March	2578	SP				cluster_2
Row23	2022	January	2749	SP				cluster_2
Row21	2021	November	3800	SP				cluster_2
Row2	2020	April	4478	SP				cluster_2
Row20	2021	October	4485	SP				cluster_2
Row9	2020	November	5568	SP				cluster_2
Row24	2022	February	6578	SP				cluster_2
Row8	2020	October	7378	SP				cluster_2
Row19	2021	September	7759	SP				cluster_2
Row10	2020	December	9244	SP				cluster_1
Row3	2020	May	10480	SP				cluster_1
Row7	2020	September	11216	SP				cluster_1
Row11	2021	January	12634	SP				cluster_1
Row12	2021	February	12918	SP				cluster_1
Row18	2021	August	13680	SP				cluster_1
Row6	2020	August	14034	SP				cluster_1
Row4	2020	June	14296	SP				cluster_1
Row5	2020	July	16468	SP				cluster_1
Row17	2021	July	22630	SP				cluster_1
Row13	2021	March	30318	SP				cluster_0
Row15	2021	May	30366	SP				cluster_0
Row16	2021	June	32614	SP				cluster_0
Row14	2021	April	43078	SP				cluster_0

Intepretação dos Resultados

Em análise do resultado da árvore de decisão confirmou minha decisão em optar pelo estado de SP para análise dos casos de óbitos registrados, mesmo MG tendo o maior número de linhas registrados o caso de óbitos registrados foi de 114.929 que representa 35% menos do que o número total registrado em SP de 321.000.

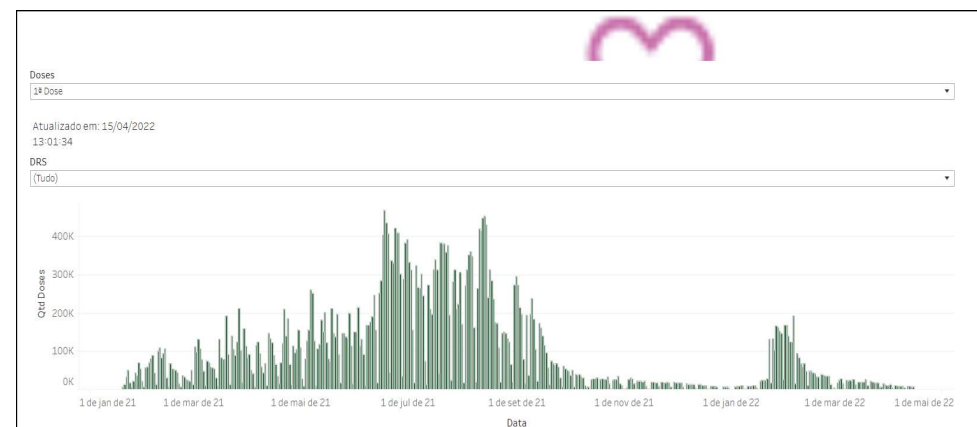
Em análise do resultado do Algoritmo Fuzzy C-Means é possível concluir que para os meses em que a campanha de Vacinação realmente atingiu o maior número de pessoas foram efetivamente os meses do qual houve a redução do número de Óbitos Registrados

Intepretação dos Resultados

Dados Categorizados no Cluster 0:

- Meses de Abril, Junho, Março, e Maio de 2021
- Estes meses são os meses que temos o maior número de registros de casos de Óbito por Covid, é possível observar que estes meses são os que estão registrados o maior número de casos por Covid portanto a questão do aumento do número de óbitos é consequência deste aumento dos casos.

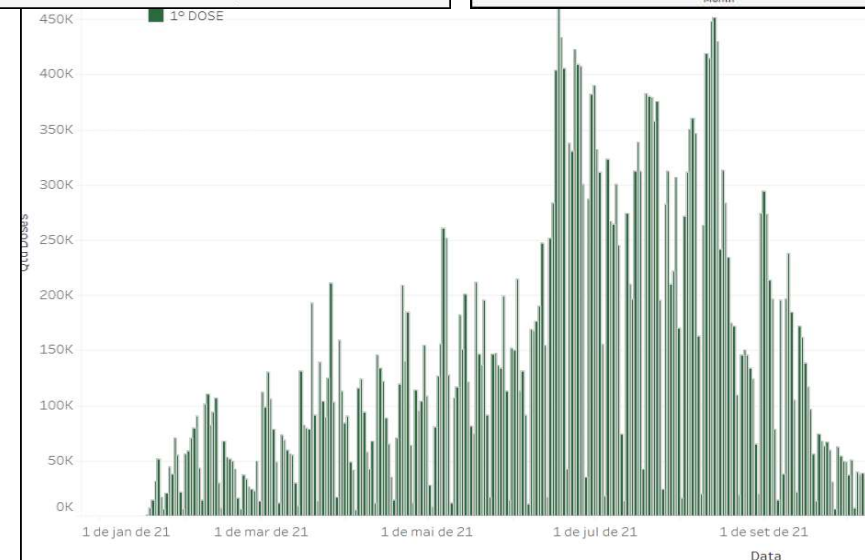
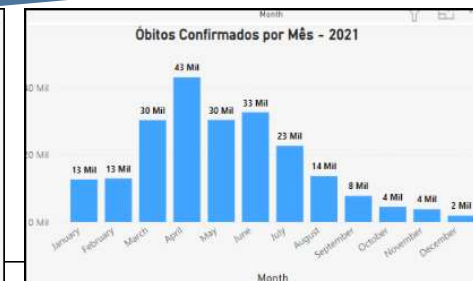
Nestes meses já estávamos vivendo novas “Ondas” e novas variantes da doença surgiram, um momento bastante delicado para toda a população. Neste momento a campanha de vacinação estava sendo disponibilizada para pessoas idosas, profissionais da área da saúde e pessoas com comorbidades, não atingindo ainda a maior parte da população do estado



Intepretação dos Resultados

Dados categorizados – Cluster 1

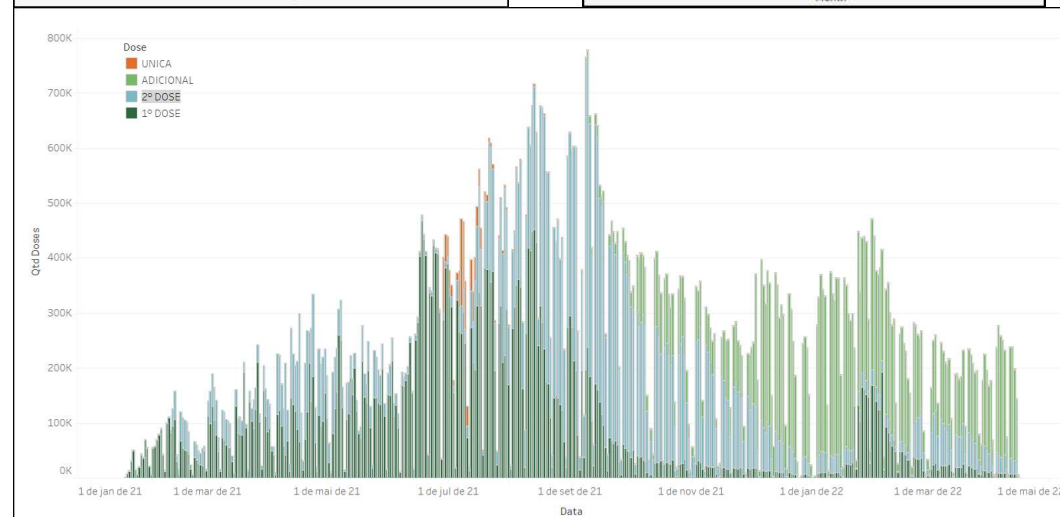
- Meses de Maio, Junho, Julho, Agosto, Setembro, Dezembro (2020)
- Meses de Janeiro, Fevereiro, Julho e Agosto (2021)
- Estes meses são os meses que temos uma variedade de cenários
- Temos registros de meses de 2020 do qual ainda não tínhamos acesso a vacina e estávamos no início da Pandemia com um aumento crescente no número de casos.
- Temos registros de meses de 2021 do qual em Janeiro e Fevereiro tivemos o início da aplicação de vacina para os idosos e grupos de risco, porém ainda não atingindo a maior parte da população
- Para os meses de Julho e Agosto estes foram os meses do qual foi iniciado a campanha de vacinação para a população adulta com os picos de vacinação no dia 13/08 atingindo 451.822 pessoas
- Inclusive fica claro de que o mês de Agosto que atingiu o pico de vacinação foi o mês que efetivamente começou a reduzir o número de casos confirmados.
- De 611mil de Julho/21 para 410mil em Agosto/21 e reduzindo o número de óbitos de 23 mil em Julho para 14 mil em Agosto
- Este gráfico já demonstra claramente a efetividade da vacina na redução dos números



Intepretação dos Resultados

Dados categorizados – Cluster 2

- Meses de Fevereiro, Março, Abril, Outubro e Novembro de 2020
 - Meses de Setembro, Outubro, Novembro e Dezembro de 2021
 - Meses de Fevereiro, Março, Abril de 2022
- Temos registros de meses de 2020 do qual ainda tínhamos um número baixo registrado pois era o momento que o Covid chegava em nosso país
- Os meses categorizados em 2021 já é observado uma queda no número de casos confirmados e consequentemente óbitos pois a aplicação da segunda dose também chegava para acesso a população tendo seu pico em 09/09/2021 sendo aplicada em 530.189 pessoas.
- Os meses categorizados em 2022 já começamos a ver medidas de restrição sendo retiradas uma vez que houve uma drástica queda no número de casos, este cluster também confirma como a aplicação da campanha de vacinação realmente foi efetiva para a contenção do vírus



Apresentação dos Resultados

Portanto em conclusão, ficou claro de que com todos os dados aqui demonstrados com o avanço da campanha de vacinação os números referentes aos casos de covid confirmados e principalmente os casos de óbitos confirmados foram reduzidos.

Ao decorrer do trabalho foi construído uma análise de contagem de dados no Python, um dashboard no Power Bi para visualização gráfica e análise de dados e utilizado dois modelos de Machine learning Árvore de Decisão e o algoritmo Fuzzy C-Means sendo este o que apresentou os melhores resultados e colaborou com a análise e conclusão da pesquisa.

Provando assim com os dados a proposta deste trabalho, demonstrar e provar com que a vacina é sim efetiva, salva vidas além disso também proporcionar a população tomar a melhor decisão com base em uma na análise de dados demonstrados.

Também efetuei o modelo de Vasandani para demonstrar a proposta e resultado atingido com a evolução deste trabalho.

Data Science Workflow Canvas*

Start here. The sections below are ordered intentionally to make you state your goals first, followed by steps to achieve those goals. You're allowed to switch orders of these steps!

1 Problem Statement What problem are you trying to solve? What larger issues do the problem address? Este problema é importante para conscientizar o cidadão da importância da Vacina contra a Covid. A proposta é provar com dados a redução do número de Óbitos causados por Covid após a campanha de vacinação em massa, para que com dados o cidadão possa ter uma visão que comprove a eficácia da vacina.	2 Outcomes/Predictions What prediction(s) are you trying to make? Identify applicable predictor (x) and/or target (y) variables. O principal resultado será a redução de casos de Óbito por covid a partir da campanha massiva de vacinação. A campanha de vacinação iniciou no Brasil em 17/02/21 a variável perceptível é que o número começará a reduzir a partir de Agosto pois a partir de 30/07/2021 a vacina foi disponibilizada para a população acima de 30 anos portanto atingindo a maior parte da população e consequentemente reduzindo os números de Óbitos por covid.	3 Data Acquisition Where are you sourcing your data from? Is there enough data? Can you work with it? O Dataset utilizado para o desenvolvimento desta pesquisa foi obtido através da página Brasil.io , página do qual tem por missão tornar acessível os dados brasileiros de interesse público e que tem como valor principal a transparência e colaboração. Os dados são suficientes para a solução do problema proposto neste trabalho.
4 Modeling What models are appropriate to use given your outcomes? O dataset utilizado possui dados rotulados e contínuos. O resultado será interpretado com base na comparação da soma de quantidade de Óbitos Registrados + A inclusão por Mês e Ano para análise comparativa.	5 Model Evaluation How can you evaluate your model's performance? A métrica avaliada será o resultado da soma dos casos de Óbitos registrados + A inclusão por mês e ano para análise comparativa.	6 Data Preparation What do you need to do to your data in order to run your model and achieve your outcomes? Exploração dos dados , analisando o total de linhas, o total de casos somados para definir um foco de análise por estado. Avaliação de dados em branco e/ou duplicados Após definição do estado a ser analisado , criar dashboards para iniciar a análise com os campos específicos da tabela , efetuando soma e acrescentando na visão Mês + Ano para análise e comparação dos resultados..

✓ Activation

When you finish filling out the canvas above, now you can begin implementing your data science workflow in roughly this order:

1 Problem Statement → 2 Data Acquisition → 3 Data Prep → 4 Modeling → 5 Outcomes/Preds → 6 Model Eval

* Note: This canvas is intended to be used as a starting point for your data science projects. Data science workflows are typically nonlinear.

“

FIM

”

IVY DE MORAES TENTONI