

# Atividade 03

03 de Novembro 2020

## 1 Tema

**Título:** Criação de índice de prioridade de vacinação das cidades em relação ao Covid-19

**Descrição:** Proposição de prioridade da distribuição de vacinas de acordo com a vulnerabilidade de cidades por meio da análise de fatores como contaminação, IDH, índices econômicos e demográficos.

## 2 Equipe

Nome da equipe: DICE-Science

Integrantes:

- Otávio Thomas Bertucini, 2023970, otaviobertuccini, otaviobertucini@gmail.com, bacharelado em Sistemas de Informação, UTFPR
- Luan Carlos Klein, 2022613, luancklein, luanklein@alunos.utfpr.edu.br, bacharelado em Engenharia de Computação, UTFPR
- Gabriel Rauta Buiar, 2022532, gabrielbuiar, gabrielbuiar@gmail.com, bacharelado em Engenharia de Computação, UTFPR

## 3 Perguntas de pesquisa e hipóteses

- Quais cidades apresentam maior vulnerabilidade em termos dos impactos na saúde referente ao coronavírus?

**Hipótese:** É possível definir a vulnerabilidade das cidades ao coronavírus através de um índice que abranja os principais indicadores sociodemográficos e de saúde referentes a cada município.

- Quais fatores podem influenciar nessa vulnerabilidade: IDH, densidade demográfica, PIB, renda per capita, quantidade de respiradores disponíveis, IDsus etc?

**Hipótese 1:** Quanto melhores os índices demográficos apresentados pelas cidades, menor a tendência dela ser mais vulnerável ao COVID-19, sofrendo

menos em termos de mortes/internações/casos entre outros aspectos relacionados à doença.

**Hipótese 2:** Quanto melhor a estrutura sanitária da cidade, menor a vulnerabilidade ao coronavírus

- É possível sugerir um índice que visa a facilitação na distribuição efetiva da vacina?

**Hipótese:** Sim. Ao realizar a análise dos dados citados acima, juntamente com uma combinação dos fatores agravantes, acredita-se que é possível gerar um índice que indique quais cidades são mais vulneráveis a sofrerem consequências do coronavírus, sugerindo assim, um indicador que fornece apoio a distribuição das vacinas nas cidades brasileiras.

## 4 Dados e modelos

### 4.1 Dados

**Dados do IDSUS:** O IDSUS é um conjunto de indicadores que visa aferir a qualidade do SUS (Sistema Único de Saúde) nas cidades brasileiras. Para esse trabalho, foi considerado o indicador de cobertura populacional, que mede a cobertura das equipes básicas de saúde. Por meio da relação desse índice com o número de casos por 100 mil habitantes e número de mortes absolutas por cidade, foi possível observar que municípios com menor cobertura médica foram mais afetados pelo vírus.

**Dados IVC:** O Índice de Vulnerabilidade Municipal ao alastramento do Coronavírus (IVC) é um índice que utiliza em seu cálculo dados demográficos, e que foi desenvolvido no início da pandemia no Brasil. O cálculo dele, como demais dados e informações está disponível no site:  
<https://fpabramo.org.br/2020/04/16/estudo-ranqueia-municipios-mais-vulneraveis-ao-coronavirus>.

**Dados Censo:** Os dados do Censo que foram utilizados foram do Censo de 2010 e são principalmente os dados demográficos: IDH, PIB, População e Renda per Capita de cada cidade.

**Dados Ocupação hospitalar:** Os dados de ocupação hospitalar foram disponibilizados pelo ministério da saúde através do link:  
<https://dados.gov.br/dataset/registro-de-ocupacao-hospitalar/resource/51f702a4-45f2-437d-9345-d7f93c2f1342>. Esse conjunto de dados mostra a quantidade de leitos específicos para SRAG, leitos normais e leitos de UTI e número de respiradores.

**Dados Covid-19:** Esse *dataset* faz referência aos dados gerais do coronavírus nas cidades brasileiras, tais como casos totais de infecção, quantidade

de mortes, taxa de mortes e casos por 100 mil habitantes. Os dados tem granularidade municipal, incluindo todas as cidades brasileiras, além de serem atualizados diariamente. Eles foram coletados através da plataforma *brasil.io*.

## 5 Modelos

**K-means:** Unindo os dados de internações (taxas de ocupação de leitos), os dados gerais do Covid (casos por 100 mil habitantes, taxas de mortes), os dados do IVC, é possível utilizar o algoritmo K-means para realizar uma clusterização das cidades, dividindo-as em grupos parecidos dentro dos termos citados acima. Essa divisão será útil para auxiliar na construção do índices, uma vez que ela já identifica grupos de cidades que apresentam uma maior vulnerabilidade. Vale destacar que essa clusterização é limitada aos índices apresentados, não valendo assim como um resultado absoluto e definitivo.

A Figura 1 apresenta uma divisão através do K-means em 5 clusters, utilizando o índice IVC e o número de contaminados em cada 100 mil habitantes. Nele se apresentam, teoricamente, a vulnerabilidade das cidades. A ordem de vulnerabilidade se daria pelas cores amarelo, azul, roxo, verde-limão e verde, em ordem decrescente.

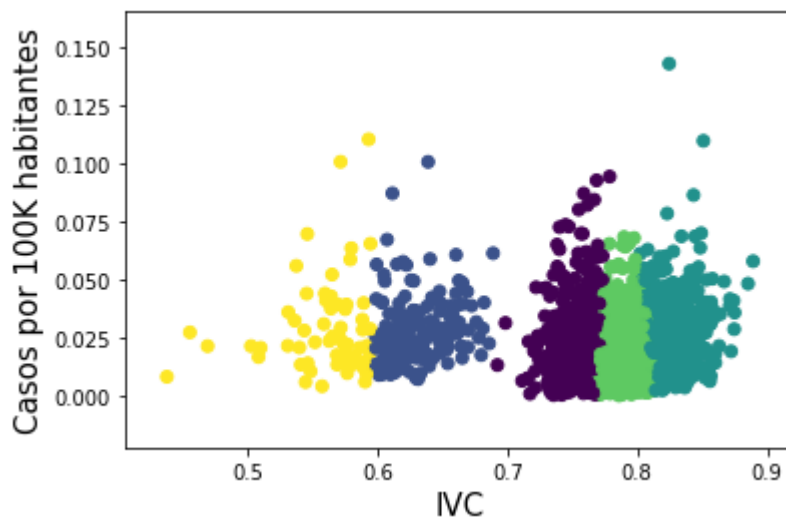


Figure 1: Exemplo de clusterização com 5 clusters, utilizando o índice IVC e o número de casos por 100 mil habitantes

**PCA:** Pretende-se usar o PCA para reduzir a dimensionalidade dos dados

de internação do SUS, removendo dados com pouca relevância do conjunto de dados.

**Local Outlier Factor (LOF):** pretende-se utilizar esse método para detectar *outliers* no conjunto de dados e cidades que precisarão de mais atenção nas campanhas de vacinação.

## 6 Cronograma

- Detecção de outliers - Luan e Otávio - (17/11/2020);
- Desenvolvimento do PCA com dados do SUS - Luan e Gabriel - (19/11/2020);
- Aprimoramento da utilização do K-means, adaptando-o para a utilização na criação do índice para a classificação das cidades - Gabriel, Luan, Otávio - (21/11/2020);