

TECH CHALLENGE 3

BIG DATA

Leonardo Cesar de Lima – RM 360109

Pós Tech – Data Analytics

FIAP

Introdução da Análise – Tech Challenge

Como parte de um Tech Challenge em Data Analytics (Big Data), este projeto tem como objetivo principal desenvolver uma análise estratégica utilizando dados públicos da PNAD COVID-19, elaborada pelo IBGE, com o intuito de subsidiar a tomada de decisão de um grande hospital frente a um possível novo surto da COVID-19.

Fonte dos dados: <https://covid19.ibge.gov.br/pnad-covid/>

A análise foi desenvolvida na plataforma Google Colab, com suporte de bibliotecas como Matplotlib, Pandas, PySpark.sql, entre outras, conforme detalhado na sessão Importando Bibliotecas do notebook.

O estudo considera 20 variáveis selecionadas da base original da PNAD, abrangendo os meses de maio, junho e julho de 2020. A seleção das variáveis foi guiada por critérios de relevância analítica, cobertura populacional e impacto no contexto clínico e socioeconômico.

A análise está organizada em três eixos temáticos principais:

1. Características clínicas dos sintomas
2. Comportamento da população durante o período pandêmico
3. Condições econômicas e sociais da população

O objetivo central é identificar pontos críticos e oportunidades de melhoria em informações que possam contribuir para:

- O reconhecimento precoce de sintomas e definição de perfis de risco;
- A identificação de vulnerabilidades sociais e econômicas por estado;
- O planejamento estratégico de recursos médicos e sociais com base em dados demográficos e comportamentais.

Nos tópicos a seguir, são apresentados os principais resultados da análise, acompanhados por gráficos, tabelas e interpretações críticas, que visam fornecer subsídios para um plano de resposta mais eficiente, baseado em evidências, diante de potenciais cenários futuros de crise sanitária.

O gráfico abaixo apresenta a quantidade de pessoas entrevistadas em cada estado durante o período analisado.

Esses dados permitem visualizar a distribuição da amostra por UF e identificar quais estados concentraram o maior e o menor número de registros de pessoas contaminadas pela COVID-19, contribuindo para uma análise comparativa regional sobre a incidência da doença.

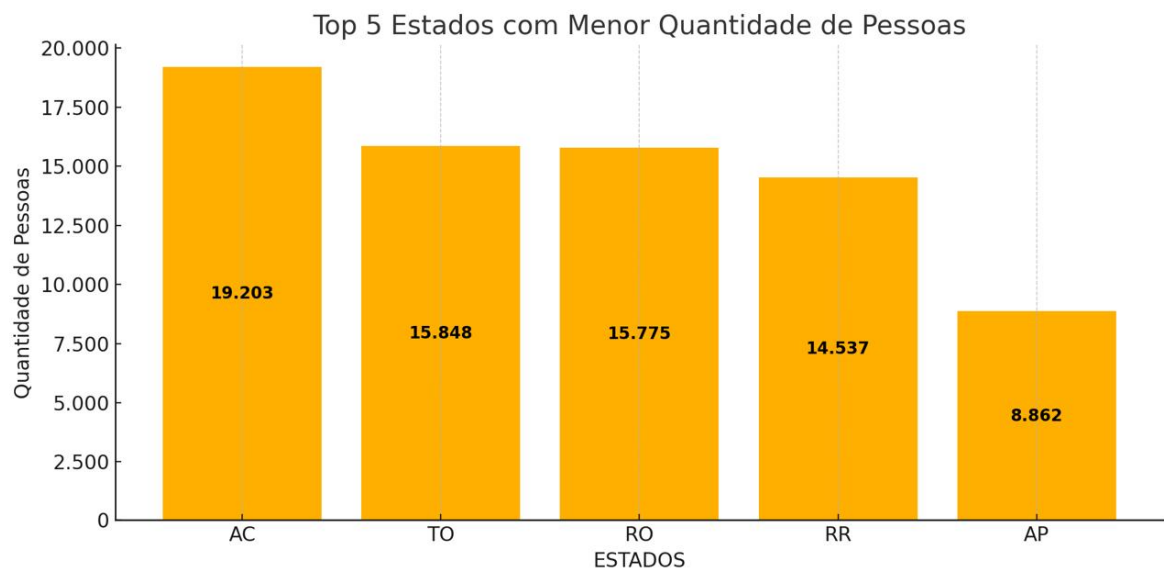


Os dois gráficos a seguir apresentam, respectivamente, os cinco estados com maior e menor proporção de pessoas contaminadas pela COVID-19.

Observa-se que os estados da Região Sudeste concentram os maiores volumes de casos, o que pode ser atribuído à alta densidade populacional, à intensa movimentação urbana e à presença de grandes centros metropolitanos, fatores que favorecem a disseminação do vírus.

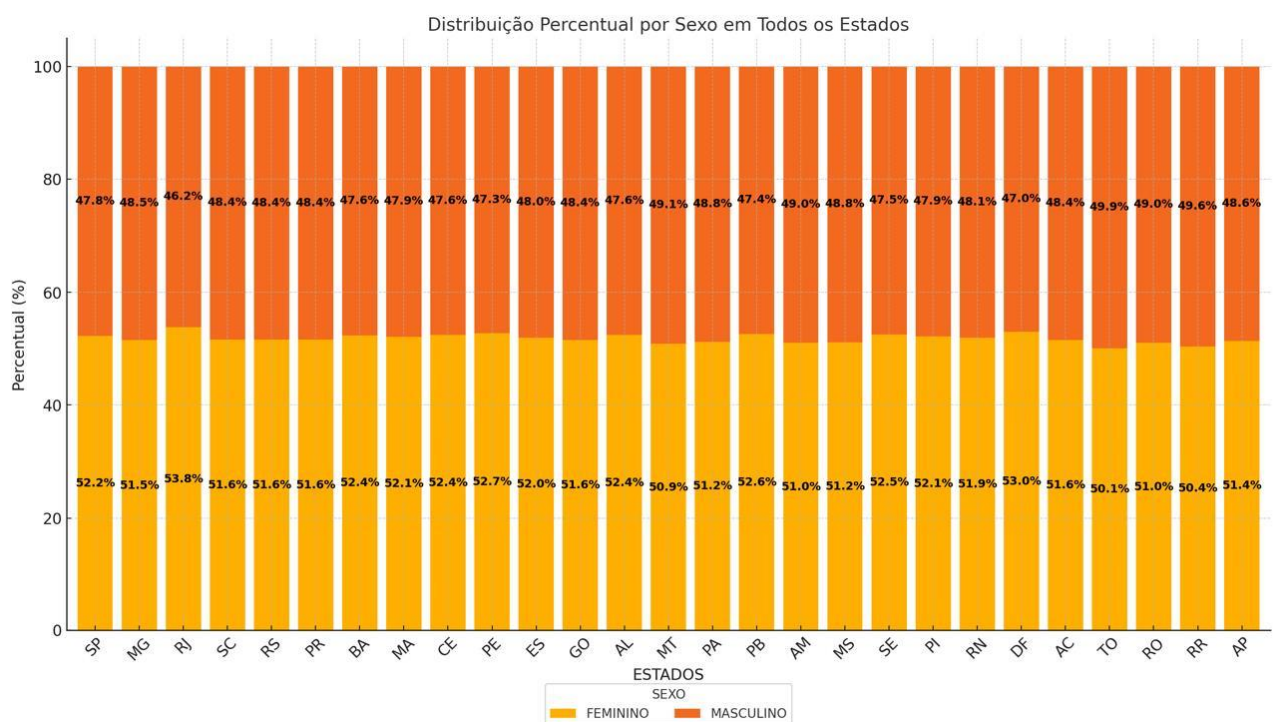


Por outro lado, os estados com menores índices estão majoritariamente localizados na Região Norte, onde características como menor concentração populacional, dispersão geográfica e infraestrutura urbana mais limitada podem ter influenciado tanto a dinâmica de contágio quanto o volume de registros coletados.



Distribuição por Sexo – Características Sociais

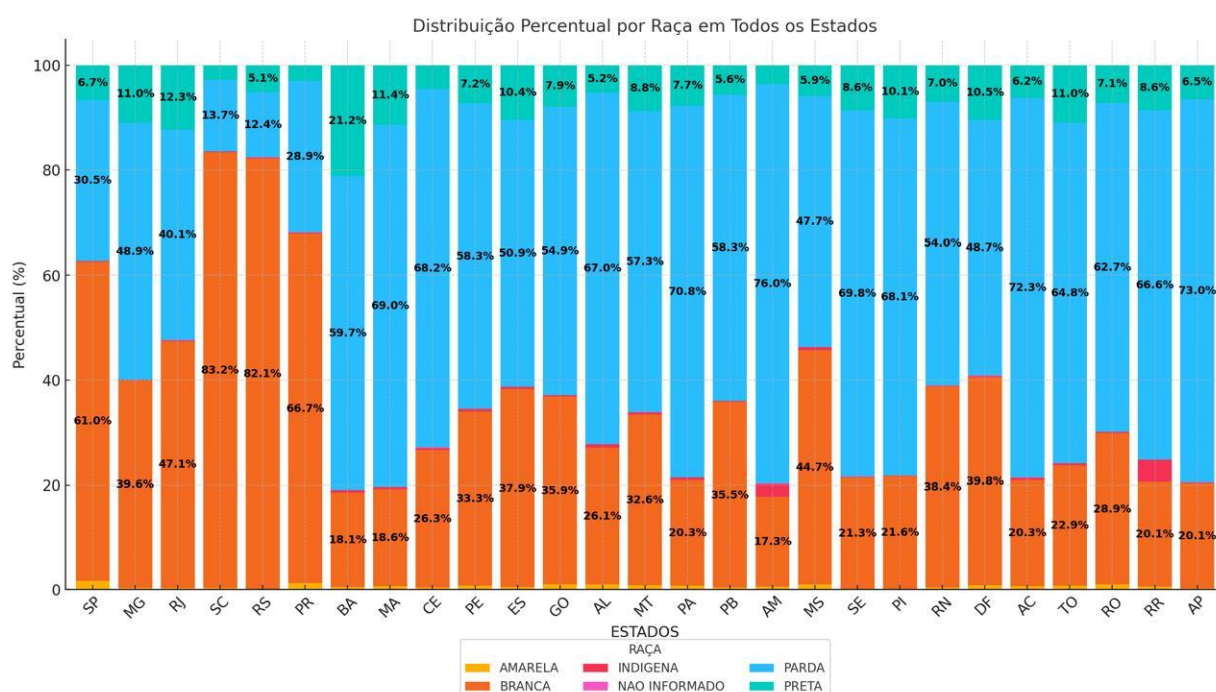
A distribuição por sexo entre os entrevistados mostra uma representação equilibrada entre homens e mulheres. Esse equilíbrio garante maior confiabilidade à análise, permitindo observar de forma justa os impactos da COVID-19 sobre ambos os grupos, tanto em aspectos clínicos quanto sociais e econômicos.



Distribuição por Raça – Características Regionais

A análise da distribuição por raça indica que os grupos branco e pardo foram os mais afetados durante o período da COVID-19. Observa-se que a população branca predominou entre os registros das regiões Sul e Sudeste, refletindo a composição demográfica dessas áreas. Já a população parda teve maior representatividade nas demais regiões do país, especialmente no Norte e Nordeste, onde compõe a maior parte da população.

Essa distribuição reforça a importância de considerar contextos regionais e sociais na formulação de políticas de saúde, uma vez que o impacto da pandemia se manifesta de forma desigual entre os diferentes grupos raciais e geográficos.

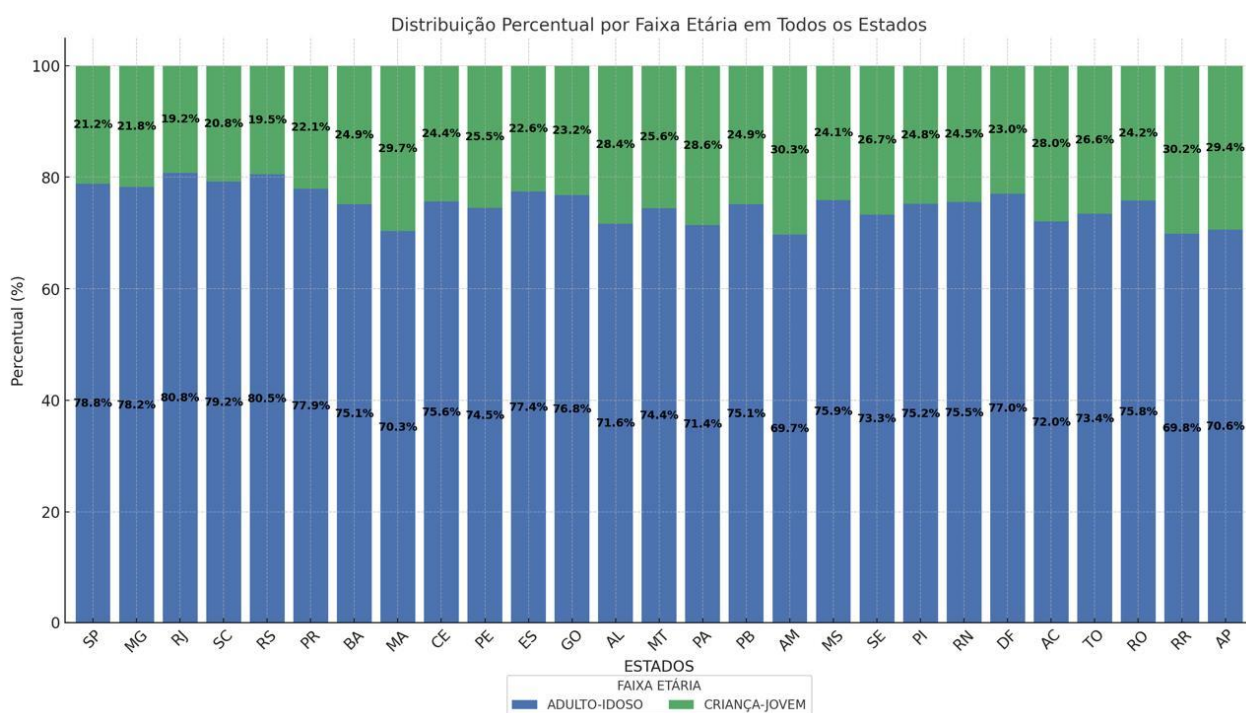


Distribuição por Faixa Etária – Impacto da COVID-19

A análise da faixa etária mostra que os adultos e idosos foram os grupos mais impactados pela COVID-19, enquanto crianças e adolescentes apresentaram menor representatividade entre os contaminados. Esse padrão pode estar relacionado a diversos fatores, como maior exposição dos adultos ao ambiente externo por motivos de trabalho e responsabilidades familiares, além da maior vulnerabilidade dos idosos a complicações da doença.

Já a menor incidência entre crianças e adolescentes pode refletir tanto um menor risco de contágio e agravamento clínico, quanto um possível menor acesso a testagens ou subnotificação nessa faixa.

Esses dados reforçam a necessidade de estratégias específicas de proteção para a população economicamente ativa e de atenção redobrada ao grupo idoso em contextos de crise sanitária.

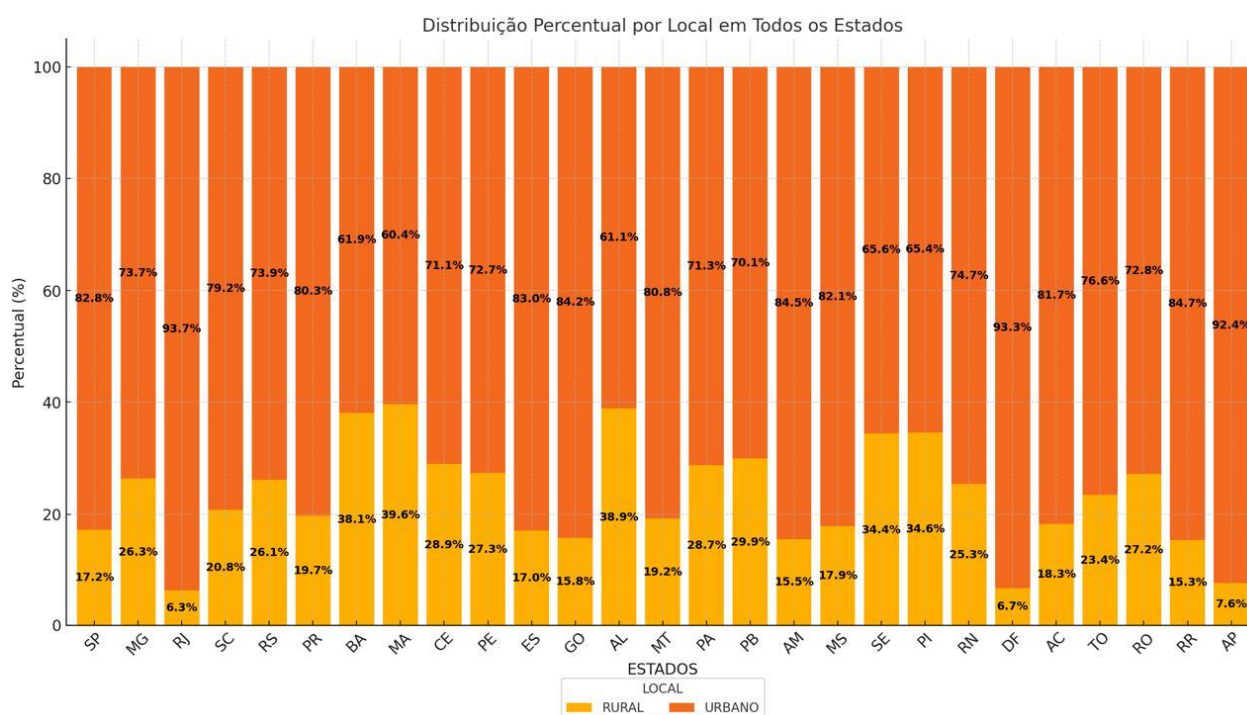


Distribuição por Local de Domicílio – Urbano e Rural

A análise da localização dos domicílios revela que a maior parte dos casos de contaminação por COVID-19 ocorreu em áreas urbanas, especialmente nos estados mais urbanizados do país, como os das Regiões Sudeste e Sul. Isso está relacionado à maior densidade populacional, maior circulação de pessoas e intensa atividade econômica, fatores que favorecem a disseminação do vírus.

Entretanto, chama a atenção o fato de que, em alguns estados do Nordeste, houve um aumento significativo de casos em áreas rurais. Esse comportamento pode indicar dificuldades no acesso à informação, à testagem e aos serviços de saúde, além da possível interiorização do vírus após os picos nos grandes centros urbanos.

Esses dados destacam a importância de políticas de saúde pública adaptadas à realidade de cada território, incluindo ações específicas para populações rurais em cenários de pandemia.

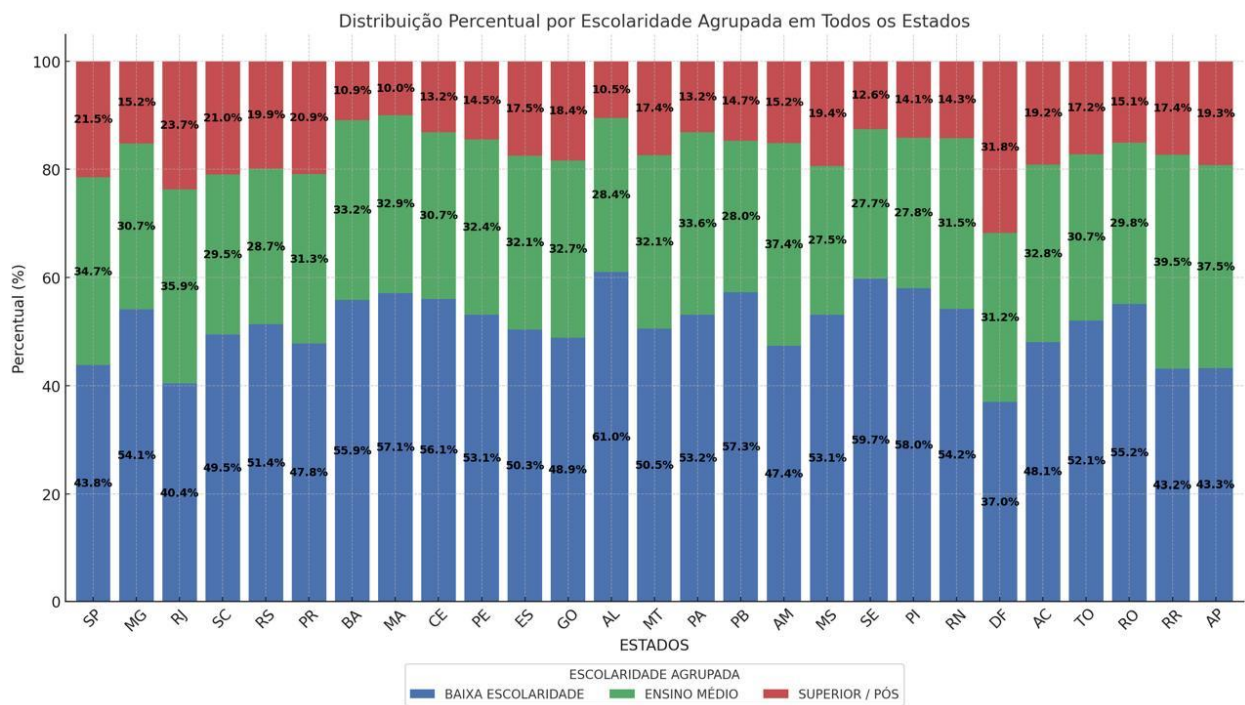


Distribuição por Escolaridade – Perfil da População Contaminada

A análise dos dados revela que a maior parte da população contaminada pela COVID-19 possuía baixa escolaridade. Em mais de 50% dos casos, os indivíduos tinham apenas o ensino fundamental completo ou incompleto, o que reforça a predominância de um público com menor nível de instrução formal.

Quando ampliamos a análise para incluir também o ensino médio, observamos que entre 80% e 90% da população contaminada possuía escolaridade até o nível médio. Por outro lado, o grupo com ensino superior, mestrado ou doutorado representou uma parcela bastante reduzida da base, estando presente em poucos estados com volume relevante.

Esse perfil reforça a necessidade de estratégias de comunicação e prevenção voltadas a públicos com menor acesso à educação formal, além de destacar uma possível relação entre nível de escolaridade e vulnerabilidade social, especialmente no contexto de crise sanitária.



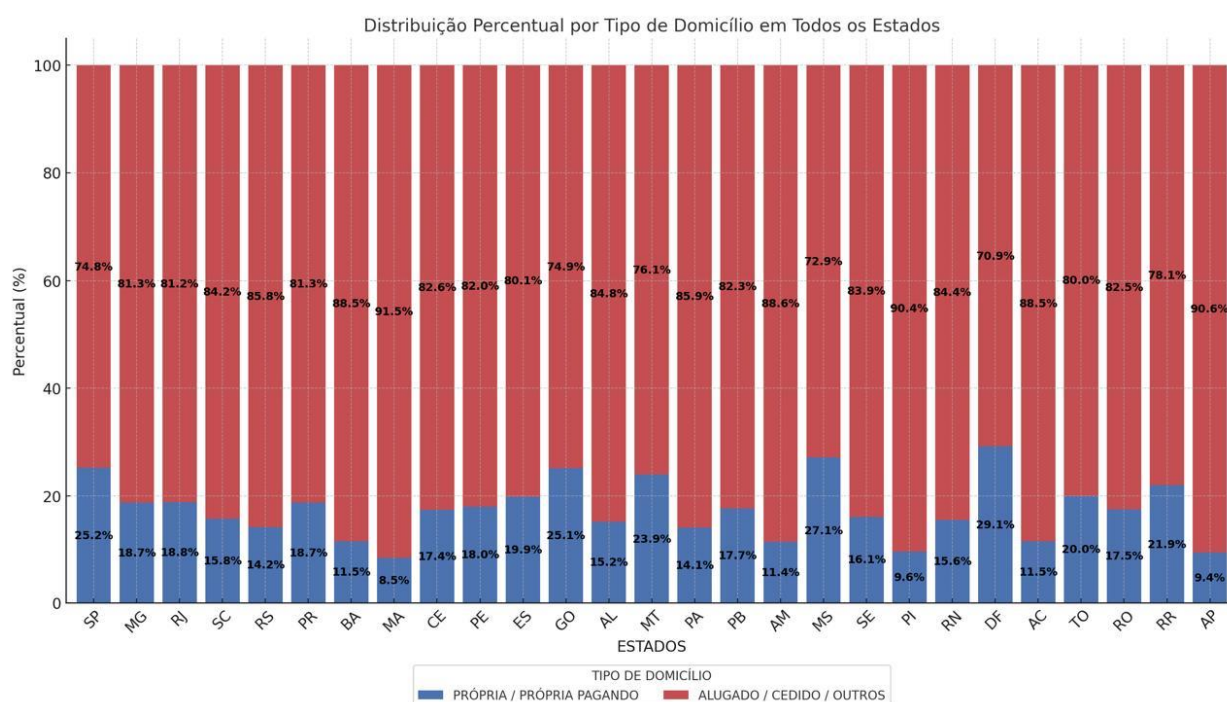
Condições de Moradia – Situação Habitacional da População Contaminada

A análise da variável relacionada ao tipo de domicílio mostra que a maioria da população contaminada vivia em moradias alugadas ou cedidas, refletindo um perfil de maior vulnerabilidade social. Em contraste, apenas uma parcela entre 10% e 30% das pessoas, a depender do estado, declarou residir em imóvel próprio.

Esse cenário se destaca principalmente na Região Nordeste, onde os registros de contaminação são mais altos entre moradores de habitações não próprias, possivelmente indicando maior densidade domiciliar, precariedade habitacional e menor acesso a condições adequadas de isolamento.

Já nas regiões Sudeste e Centro-Oeste, observou-se uma proporção menor de casos entre a população com moradia cedida ou alugada, sugerindo um possível impacto atenuado em áreas com melhor infraestrutura urbana e condições socioeconômicas mais estáveis.

Esses dados reforçam a importância de considerar a condição de moradia como fator de risco em situações de emergência sanitária e no planejamento de políticas públicas de saúde e assistência social.



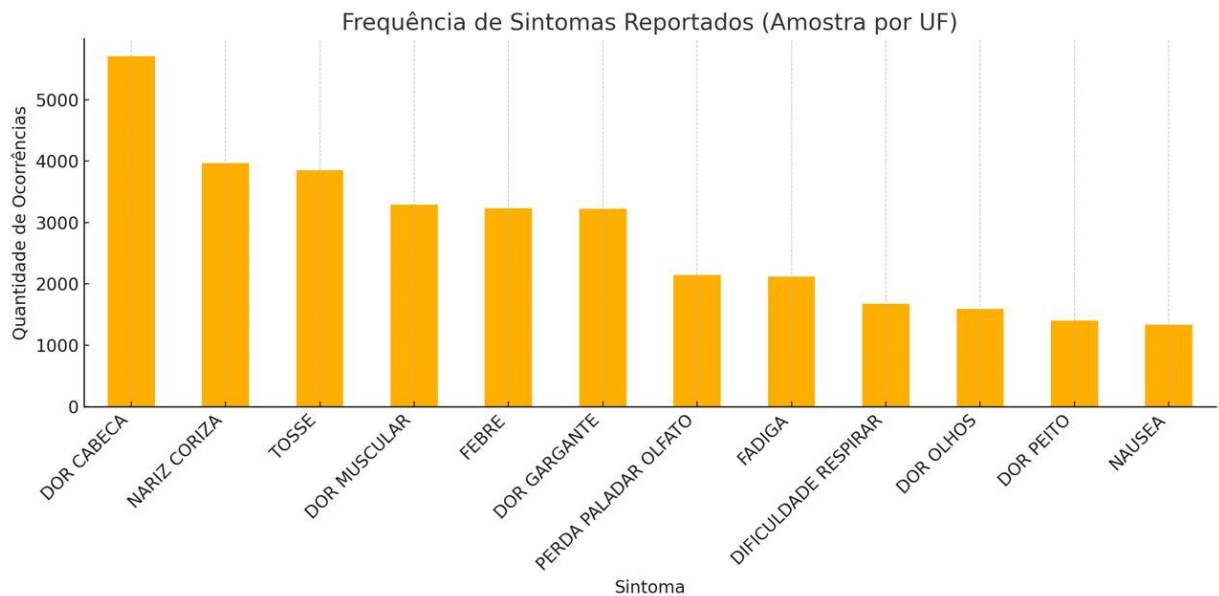
Principais Sintomas da COVID-19 – Perfil Clínico da População

A análise por amostragem dos estados revelou os principais sintomas relatados pelas pessoas contaminadas pela COVID-19, permitindo traçar um perfil clínico da população durante o período pandêmico. A frequência dos sintomas reforça os achados médicos conhecidos sobre a doença, com alguns destaques importantes observados nos dados.

Os sintomas mais recorrentes, em ordem de prevalência, foram:

1. Dor de cabeça – o sintoma mais frequentemente relatado, presente em grande parte dos casos.
2. Nariz escorrendo ou coriza – também muito comum, indicando forte impacto respiratório.
3. Tosse – outro sintoma clássico e amplamente relatado nos casos positivos.
4. Dor muscular – sinal frequente de infecção viral e presente de forma significativa na população.
5. Febre – um dos principais marcadores clínicos para identificação da COVID-19.
6. Dor de garganta – sintoma comum nos estágios iniciais da infecção.
7. Perda de paladar e olfato – sinal característico da COVID-19, relatado com frequência média nos estados.
8. Fadiga – sensação de cansaço intenso, presente em parte significativa dos casos.
9. Dificuldade de respirar – sintoma mais grave, associado a complicações respiratórias.
10. Dor nos olhos, dor no peito e náusea – menos frequentes, mas ainda observados em uma parcela dos pacientes.

Essa distribuição de sintomas é consistente com as evidências clínicas sobre a COVID-19, destacando a necessidade de atenção especial aos primeiros sinais, como dor de cabeça, tosse e febre, para garantir identificação precoce, isolamento e tratamento adequado. A presença de sintomas respiratórios graves, como dificuldade para respirar e dor no peito, também reforça a importância do monitoramento ativo e da triagem em ambientes clínicos.

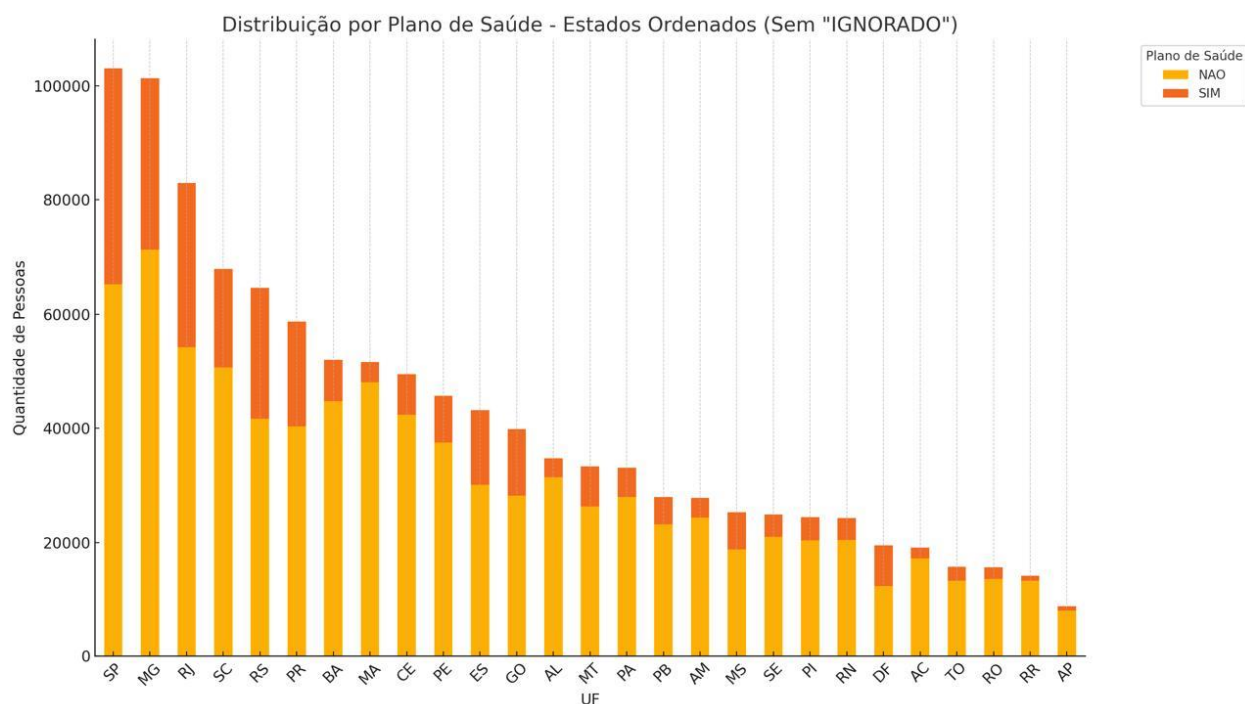


Acesso à Saúde – Diferenças Regionais no Uso de Planos de Saúde

A análise dos dados por unidade federativa evidencia uma desigualdade regional no acesso a planos de saúde no Brasil. As regiões Sul e Sudeste concentram a maior proporção da população com plano de saúde privado, o que reflete melhores condições econômicas, maior urbanização e maior oferta de serviços privados nessas áreas.

Por outro lado, nas regiões Norte e Nordeste, a maioria da população depende exclusivamente do sistema público de saúde (SUS). Esses dados evidenciam uma maior vulnerabilidade estrutural nesses territórios, tanto em termos de acesso quanto de capacidade de resposta frente a emergências sanitárias como a pandemia da COVID-19.

Essa diferença regional reforça a importância de políticas públicas direcionadas, com foco no fortalecimento da rede pública de saúde em regiões com menor cobertura privada, garantindo equidade no atendimento e na prevenção em todo o país.



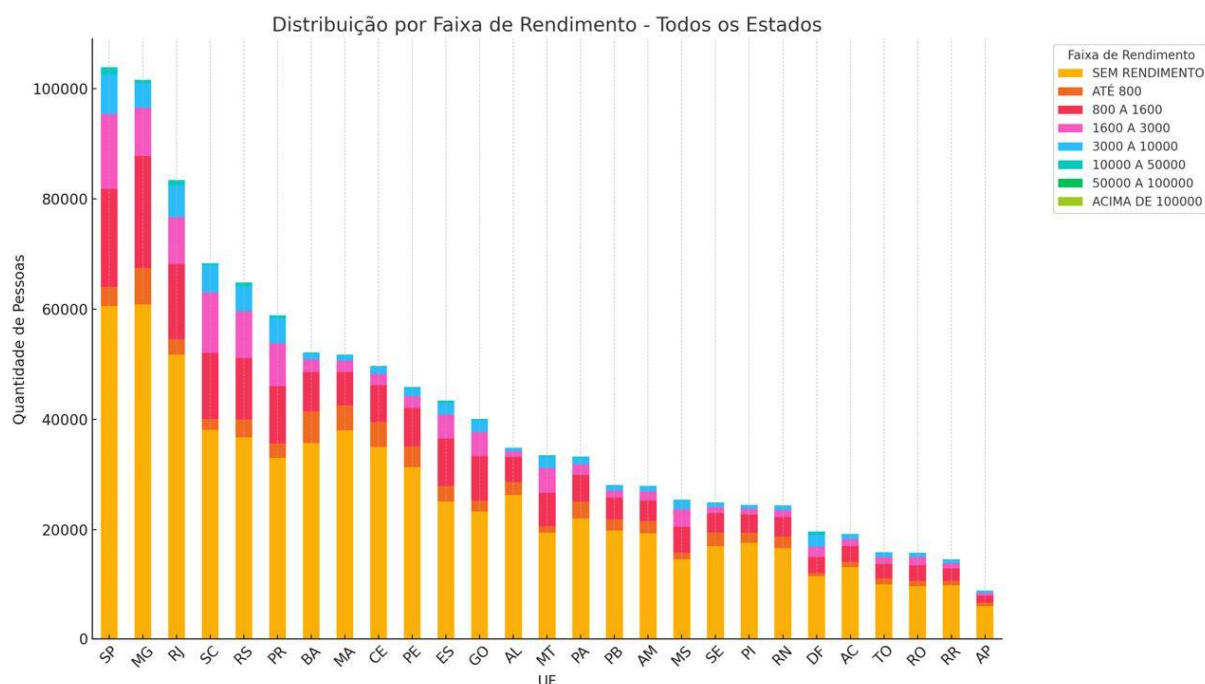
Distribuição de Rendimento – Impacto da COVID-19 entre os Grupos de Baixa Renda

A análise dos dados revela que a maior parte da população contaminada pela COVID-19 encontra-se nas faixas de renda mais baixas. A maioria dos entrevistados declarou estar sem rendimento ou receber até R\$ 1.600,00 por mês, o que indica um perfil de alta vulnerabilidade econômica.

Apenas uma pequena parcela da população, concentrada principalmente nas regiões Sul e Sudeste, declarou rendimentos entre R\$ 1.600,00 e R\$ 10.000,00. Essa diferença reflete não apenas desigualdades regionais, mas também a forma como a condição socioeconômica influenciou a exposição e o risco de contaminação.

Os dados apontam que os grupos com menor renda foram os mais afetados pela pandemia, possivelmente devido à dificuldade de realizar isolamento social, à necessidade de trabalhar presencialmente e às condições precárias de moradia e transporte.

Essa realidade destaca a importância de considerar critérios sociais e econômicos nas estratégias de prevenção e resposta à COVID-19, especialmente em contextos de crise sanitária.

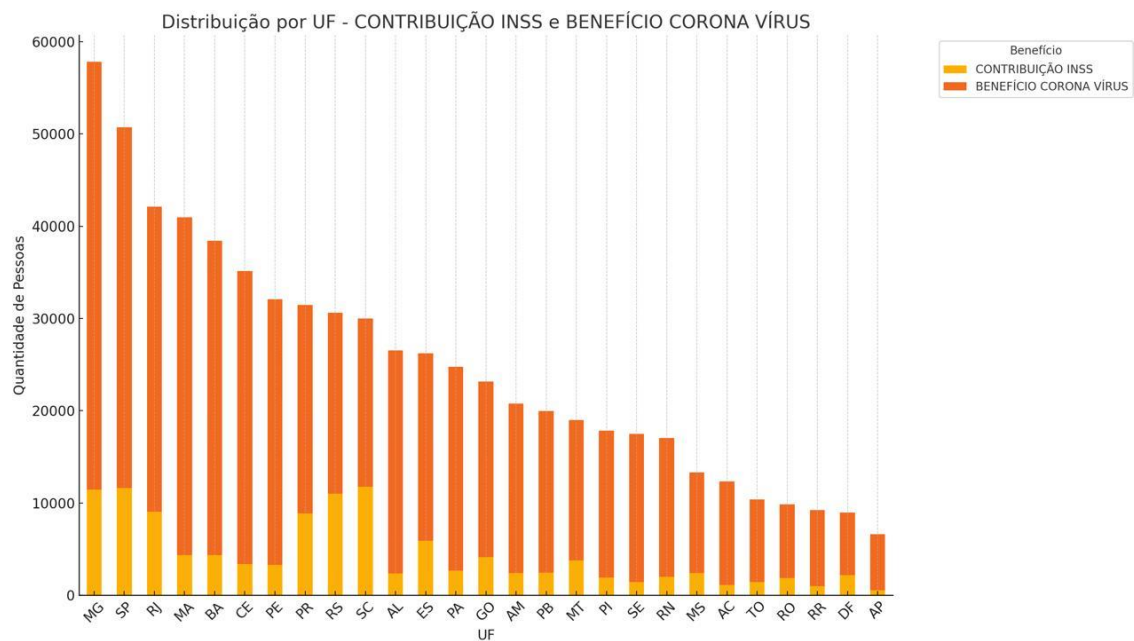
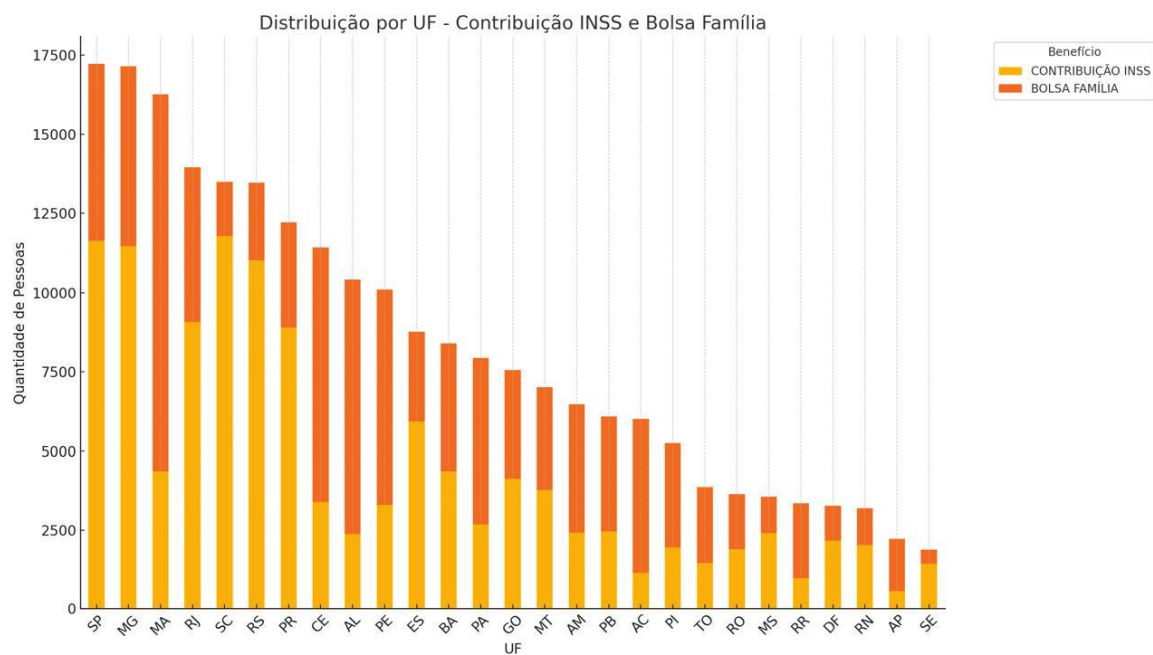


Relação entre Benefícios Sociais e Contribuição ao INSS – Desigualdade Regional

A análise dos dados evidencia que uma parcela significativa da população, especialmente nas regiões Norte e Nordeste, recebe mais benefícios sociais do que contribui para o INSS. Essa situação revela uma forte dependência de programas de assistência pública, como o Bolsa Família, BPC e Auxílio Emergencial, por parte de grupos economicamente vulneráveis.

Enquanto a contribuição ao INSS é mais frequente nas regiões Sul e Sudeste, onde há maior formalização do mercado de trabalho, nas demais regiões observa-se um baixo índice de contribuição previdenciária, associado à informalidade, ao desemprego e à baixa renda.

Esse desequilíbrio destaca a importância de políticas que incentivem a inclusão produtiva e a formalização do trabalho, ao mesmo tempo em que garantam proteção social adequada para as populações em situação de fragilidade econômica, sobretudo em regiões historicamente menos desenvolvidas.



Recomendações Estratégicas para Preparação Hospitalar diante de Nova Onda da COVID-19

Com base nos dados da PNAD-COVID e nos indicadores analisados, é possível identificar padrões claros sobre os perfis de maior risco e vulnerabilidade. A seguir, são apresentadas diretrizes fundamentais para que hospitais se preparem com antecedência para um possível novo cenário de contaminação e internações.

1. Reconhecimento Precoce de Perfis de Risco

- Monitorar sintomas prevalentes como dor de cabeça, coriza, tosse, febre e dor muscular para identificação rápida de casos suspeitos.
- Priorizar adultos e idosos, que foram os grupos mais impactados clinicamente.
- Atentar-se a pacientes com comorbidades e histórico de vulnerabilidade, como baixa escolaridade, ausência de plano de saúde ou condições habitacionais precárias.

2. Identificação de Vulnerabilidades Econômicas e Sociais

- Utilizar dados regionais para identificar populações com renda de até R\$ 1.600,00, maior incidência de contaminação e acesso limitado à saúde.

- Mapear comunidades com alto número de pessoas vivendo em moradias alugadas ou cedidas, em especial nas regiões Norte e Nordeste, onde o acesso ao sistema privado é reduzido.
- Considerar o número de benefícios sociais recebidos como um indicativo direto de fragilidade socioeconômica.

3. Planejamento Estratégico com Base na Demografia

- Desenvolver planos de ação considerando a distribuição por estado, faixa etária, sexo e raça, direcionando recursos e ações conforme as características de cada região.
- Criar protocolos diferenciados para áreas urbanas e rurais, com atenção especial à interiorização do vírus em regiões remotas.
- Priorizar a região Nordeste no planejamento de contingência, por concentrar maior número de pessoas em situação de vulnerabilidade e com menor estrutura de saúde.

4. Reforço da Capacidade de Atendimento

- Ampliar e preparar leitos clínicos e de UTI, especialmente em locais de alta densidade populacional e baixa cobertura de planos de saúde.
- Garantir estoques adequados de EPIs, testes, medicamentos e insumos respiratórios, com logística adaptada para resposta rápida.

5. Ações de Comunicação e Engajamento Comunitário

- Produzir material educativo claro e acessível, adaptado a públicos com baixo nível de escolaridade.
- Utilizar canais comunitários e mídias locais para informar sobre sintomas, prevenção e acesso ao atendimento.

6. Integração com a Rede Pública e Assistência Social

- Estabelecer fluxos integrados com postos de saúde e programas sociais, garantindo rastreabilidade e atendimento coordenado.
- Fortalecer a comunicação entre hospitais e a atenção primária para monitoramento contínuo de grupos de risco.

7. Uso de Dados para Antecipação e Decisão

- Criar painéis internos de inteligência com cruzamento de dados demográficos, clínicos e socioeconômicos.
- Apoiar decisões de contingência com base em modelos preditivos alimentados por dados públicos e institucionais, permitindo alocação mais eficiente de recursos.

Essas medidas são essenciais para garantir uma resposta hospitalar proativa, eficaz e equitativa, com base em evidências, dados regionais e compreensão profunda das realidades sociais da população brasileira.

Link GitHub: https://github.com/LeonardoRoig/TECH_CHALLENGE_3