



FACULDADE DE INFORMÁTICA E ADMINISTRAÇÃO PAULISTA

FELIPE DANTAS TAVARES
IZABEL MORAIS DE CARVALHO
JOE ALLAN ZIRN
JORGE MENEZES GUERRA MARTINS
PEDRO AUGUSTO MOREIRA

INTELIGÊNCIA DE DADOS PNAD-COVID-19

Tech Challenge

São Paulo

2026

LISTA DE ILUSTRAÇÕES

Figura 1 - Diagrama da arquitetura de solução de Big Data.....	06
Figura 2 - Dashboard: Perfil Demográfico e Populacional	08
Figura 3 - Dashboard: Jornada do Paciente e Pressão Hospitalar.....	09
Figura 4 - Dashboard: Impacto Econômico e Isolamento	10

LISTA DE TABELAS

Tabela 1 - Mapeamento das Variáveis Seleccionadas	06
---	----

SUMÁRIO

1 INTRODUÇÃO E CONTEXTUALIZAÇÃO	05
2. ARQUITETURA DE DADOS EM NUVEM (DATA LAKEHOUSE)	05
3. METODOLOGIA: MATRIZ DAS 20 VARIÁVEIS ESTRATÉGICAS	06
4. VISUALIZAÇÃO DE DADOS: DASHBOARD EXECUTIVO	08
4.1 Página 1: Perfil Demográfico e Estrutura Populacional	08
4.2 Página 2: Pressão no Sistema de Saúde	09
4.3 Página 3: Vulnerabilidade Econômica e Barreiras de Isolamento	09
5. PLANO DE AÇÃO ESTRATÉGICO	10
5.1. O "Gargalo" da Triagem	10
5.2. A Desigualdade como Vetor de Contágio	10
5.3. Fator de Risco Etário	11
6. PLANO DE AÇÃO ESTRATÉGICO	11
7. CONCLUSÃO	12
REFERÊNCIAS	13

1. Introdução e Contextualização

O presente projeto detalha a estruturação de uma plataforma de *Business Intelligence* e *Analytics* desenvolvida sob demanda para um centro hospitalar de grande porte. O objetivo central foi identificar correlações estatísticas entre escolaridade, renda, adesão ao isolamento social e desfechos clínicos.

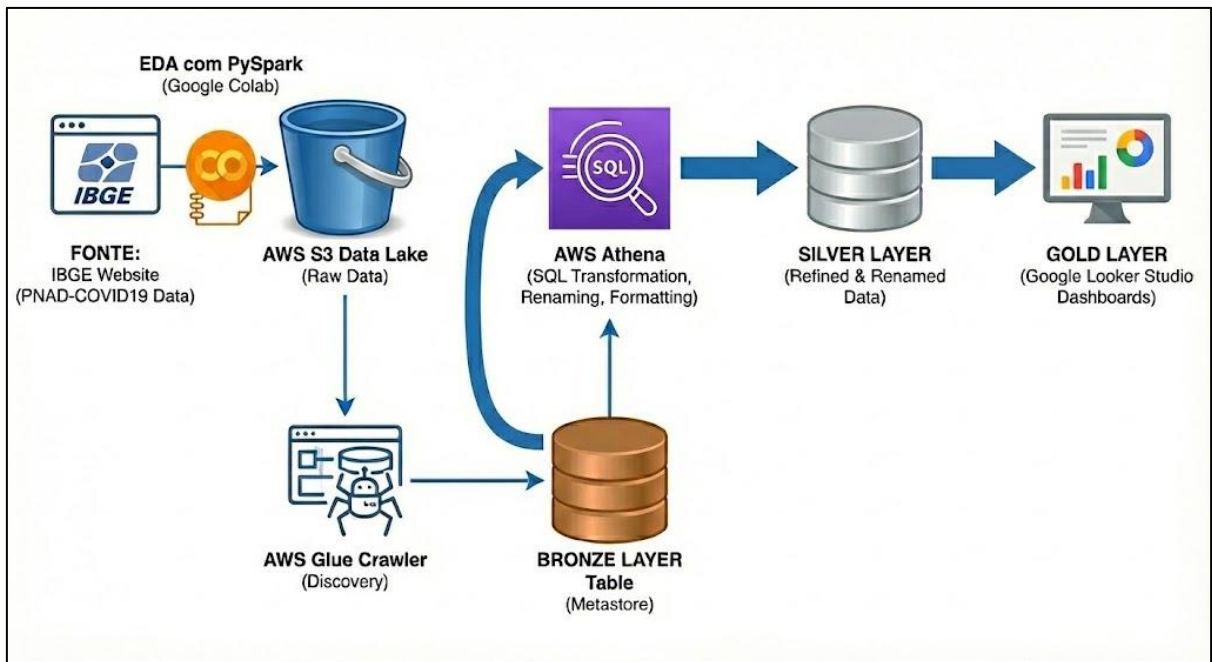
Para tal, utilizou-se a base de dados da PNAD-COVID-19 (IBGE) como "única fonte da verdade" (*Single Source of Truth*). A análise visa fundamentar o planejamento estratégico da instituição, permitindo a antecipação de demandas por leitos e insumos em cenários de novos surtos pandêmicos.

2. Arquitetura de Dados em Nuvem (*Data Lakehouse*)

Em conformidade com a solicitação da diretoria técnica, a infraestrutura foi provisionada 100% em nuvem para garantir elasticidade e alta disponibilidade. Implementou-se uma arquitetura de *Data Lakehouse* na *Amazon Web Services* (AWS), seguindo as melhores práticas da *Medallion Architecture*:

- Camada Bronze (*Raw Layer*): Responsável pela ingestão dos dados brutos no *Amazon S3*, preservando o formato original (CSV) e a integridade histórica dos meses de referência (maio, junho e julho).
- Camada Silver (*Refined Layer*): Etapa de processamento via *AWS Athena* e linguagem SQL. Nesta fase, realizaram-se tratamentos críticos: limpeza de valores nulos, correção de *schema drift* e renomeação semântica das variáveis para facilitar o entendimento de negócio.
- Camada Gold (*Analytics Layer*): Disponibilização dos dados refinados para consumo no *Google Looker Studio*, convertendo milhões de registros complexos em indicadores executivos visuais para tomada de decisão.

Figura 1: Diagrama da arquitetura de solução de Big Data



Fonte: Elaborado pelo autor (2026).

3. Metodologia: Matriz das 20 Variáveis Estratégicas

Para atender estritamente ao requisito de projeto (utilização de no máximo 20 questionamentos), realizou-se uma curadoria das variáveis, selecionando apenas aquelas com alto poder preditivo nos eixos clínico, demográfico e econômico.

A Tabela 1 abaixo detalha o dicionário de dados da solução:

Tabela 1: Mapeamento das Variáveis Selecionadas

Dimensão	Código IBGE	Justificativa Estratégica para o Hospital
Identificação	V1032	Cálculo de expansão amostral e representatividade populacional.
	V1013	Análise da evolução temporal nos meses de maio, junho e julho.
	UF	Identificação geográfica de estados com maior pressão hospitalar.
Perfil Populacional	A002	Identificação de faixas etárias de risco para priorização de leitos.

	A003	Análise demográfica e padrões de internação por gênero.
	A004	Avaliação de disparidades no acesso à saúde e vulnerabilidade social.
	A005	Cruzamento crítico entre nível de instrução e taxas de internação.
Triagem Clínica	B0011	Monitoramento de febre como sintoma sentinela para triagem.
	B0012	Monitoramento de tosse para análise de propagação viral.
	B0014	<i>KPI</i> Crítico: Principal preditivo de necessidade de UTI e oxigênio.
Uso do Sistema	B002	Medição da demanda real por busca de atendimento médico.
	B005	Determinação da taxa de ocupação hospitalar e severidade clínica.
	B009B	Validação laboratorial (SWAB) para planejamento de alas de isolamento.
Trabalho e Isolamento	C001	Identificação da exposição da população economicamente ativa.
	C002	Monitoramento da adesão ao isolamento por motivos laborais.
	C003	Detalhamento da natureza do afastamento (preventivo vs. clínico).
	C007C	Avaliação do nível de proteção laboral (vínculo formal vs. informal).
	C013	Medição da eficácia do teletrabalho como barreira sanitária.

Características Econômicas	D0051	Análise de dependência e impacto do suporte estatal na população.
	D0053	Quantificação do suporte financeiro médio para planejamento social.

Fonte: Elaborado pelo autor (2026).

4. Visualização de Dados: Dashboard Executivo

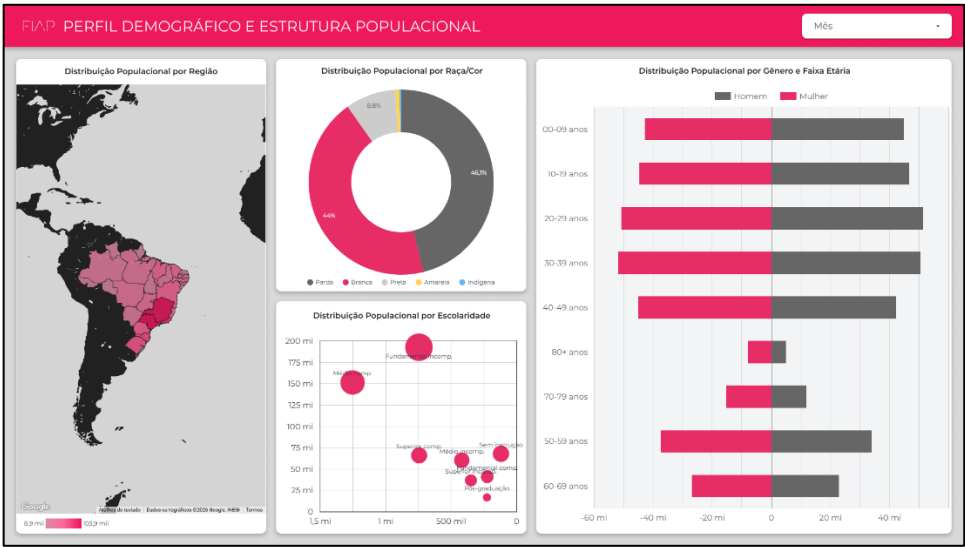
Para materializar a camada *Gold* do *Data Lakehouse*, desenvolveu-se um Painel de Inteligência no *Google Looker Studio*, estruturado em três páginas temáticas que narram a crise sanitária sob diferentes óticas: Demográfica, Clínica e Econômica.

4.1 Página 1: Perfil Demográfico e Estrutura Populacional

Esta página foca na caracterização da base populacional monitorada, permitindo entender "quem" é o paciente em potencial.

- Distribuição Regional: Um mapa de calor (*heatmap*) evidencia a densidade populacional por Estado, permitindo visualizar a concentração da demanda.
- Segmentação: O gráfico de rosca destaca a predominância das categorias Parda (46,1%) e Branca (44%), enquanto a Pirâmide Etária confirma o padrão demográfico brasileiro (base larga de jovens), fundamental para o planejamento de longo prazo.

Figura 2: Dashboard: Perfil Demográfico e Populacional



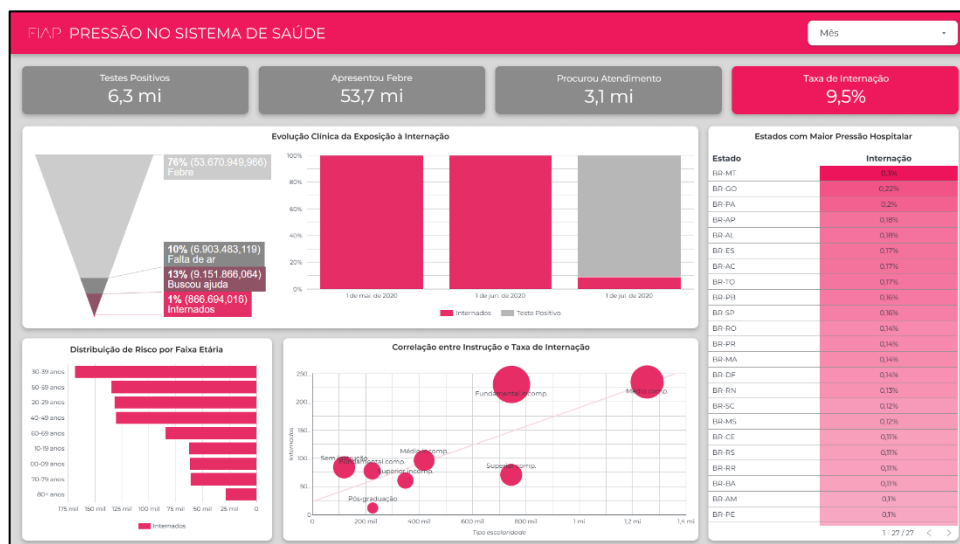
Fonte: Elaborado pelo autor (2026).

4.2 Página 2: Pressão no Sistema de Saúde

Focada na gestão hospitalar, esta visão desenha o "Funil da COVID-19", desde o sintoma até a internação.

- O Funil Clínico: O gráfico de funil ilustra a conversão crítica: de 53,7 milhões de pessoas com febre, apenas uma fração evoluiu para internação, mas a taxa de 9,5% de hospitalização (*KPI* em destaque) alerta para a necessidade de leitos.
- Gargalos: A tabela de "Estados com Maior Pressão" permite identificar quais UFs (como MT, GO, PA) estão operando próximas ao limite.

Figura 3: Dashboard: Jornada do Paciente e Pressão Hospitalar



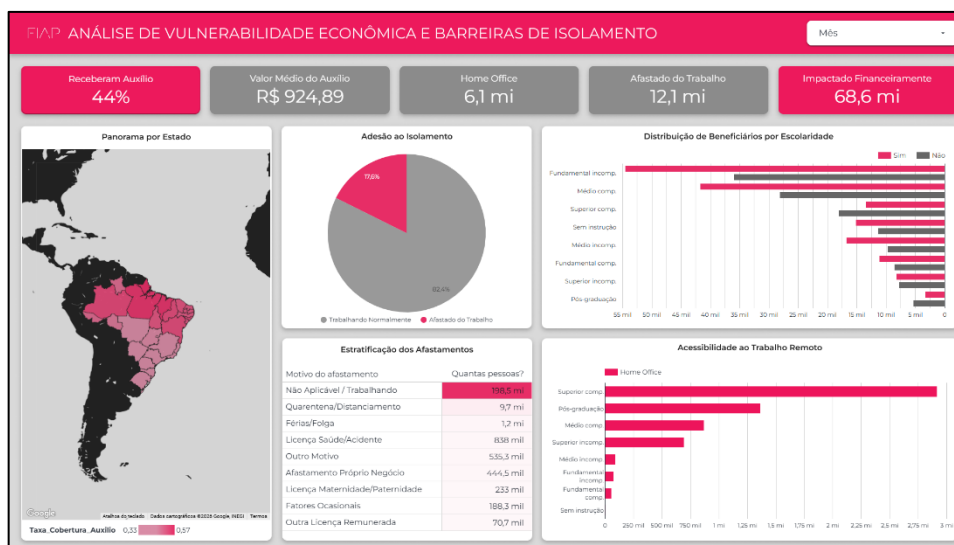
Fonte: Elaborado pelo autor (2026).

4.3 Página 3: Vulnerabilidade Econômica e Barreiras de Isolamento

Esta página cruza a saúde com a capacidade financeira de sobrevivência ao isolamento.

- A Desigualdade do *Home Office*: O gráfico de barras "Acessibilidade ao Trabalho Remoto" prova que o isolamento é um privilégio de escolaridade (Superior/Pós), enquanto a maioria da força de trabalho (12,1 milhões de afastados) depende de benefícios.
- Suporte Estatal: Os *KPIs* monitoram a cobertura do Auxílio Emergencial (44% da população) com ticket médio de R\$ 924,89, vital para reduzir a circulação de pessoas.

Figura 4: Dashboard: Impacto Econômico e Isolamento



Fonte: Elaborado pelo autor (2026).

5. Plano De Ação Estratégico

Com base nos indicadores analisados, recomenda-se à diretoria hospitalar a execução imediata das seguintes ações em caso de novo surto:

5.1. O "Gargalo" da Triagem

Observando a Figura 3, identifica-se uma discrepância comportamental grave na jornada do paciente:

- **Sintoma vs. Ação:** Embora 53,7 milhões de pessoas tenham apresentado febre (sintoma sentinela), apenas 3,1 milhões procuraram atendimento médico formal.
- **Conclusão:** Há uma subnotificação massiva na atenção primária. O paciente evita o hospital no início dos sintomas e tende a chegar apenas quando o quadro se agrava (evidenciado pelos 9,5% de taxa de internação direta entre os confirmados).

5.2. A Desigualdade como Vetor de Contágio

A análise da Figura 4 desmistifica a ideia de que o isolamento depende apenas de "conscientização":

- **Privilégio Educacional:** O gráfico de "Acessibilidade ao Trabalho Remoto" mostra que o *Home Office* foi quase nulo para pessoas com Ensino Médio ou Fundamental.

- Dependência: Com 68,6 milhões de pessoas impactadas financeiramente, a adesão ao isolamento nas classes C, D e E torna-se inviável sem o suporte do Auxílio Emergencial (R\$ 924,89). A correlação é clara: menor escolaridade = menor isolamento = maior exposição ao risco.

5.3. Fator de Risco Etário

A Pirâmide Etária (Figura 2) cruzada com o Gráfico de Risco por Faixa Etária (Figura 2) revela que, embora a base da população seja jovem, a pressão hospitalar (internações) recai desproporcionalmente sobre as faixas acima de 50 anos, validando a necessidade de segregamento de leitos.

6. Plano de Ação Estratégico

Baseado nos indicadores visuais, recomenda-se à diretoria hospitalar:

AÇÃO 1: Protocolo de Captura Precoce (Telemedicina)

- Diagnóstico: O "Funil de Sintomas" mostrou que 50 milhões de febris não buscaram o hospital.
- Ação: Implementar canal de teleatendimento para triar esses pacientes em casa, evitando que cheguem à emergência apenas quando necessitam de UTI.

AÇÃO 2: Gestão de Risco Social

- Diagnóstico: O gráfico de "*Home Office*" comprova que trabalhadores de nível médio/operacional não conseguem se isolar.
- Ação: Direcionar campanhas de distribuição de EPIs (máscaras PFF2) especificamente para este público, que é obrigado a usar transporte público e manter a economia girando.

AÇÃO 3: Alocação Regional Dinâmica

- Diagnóstico: A tabela de "Pressão por Estado" na Página 2 mostra disparidades regionais.
- Ação: Criar um comitê de crise que movimenta respiradores e equipes médicas com base no *heatmap* de internações, enviando recursos de estados em baixa para estados listados no topo do *ranking* de pressão.

7. Conclusão

A estruturação deste projeto permitiu segmentar a crise da COVID-19 em dois vetores indissociáveis: o Risco Biológico (determinado pela Idade) e o Risco Social (determinado pela Escolaridade e Renda).

Conclui-se que o sucesso institucional no enfrentamento de um novo surto não depende apenas da excelência clínica no tratamento dos idosos, mas também da compreensão de que a demanda hospitalar virá majoritariamente da parcela da população que não possui o privilégio econômico do isolamento social. A inteligência de dados aqui apresentada transforma essa percepção em estratégia acionável.

REFERÊNCIAS

AMAZON WEB SERVICES (AWS). Amazon Athena: serverless interactive query service. [S. l.]. Disponível em: <https://aws.amazon.com/athena/>. Acesso em: 27 out. 2025.

AMAZON WEB SERVICES (AWS). Amazon S3: object storage built to retrieve any amount of data from anywhere. [S. l.]. Disponível em: <https://aws.amazon.com/s3/>. Acesso em: 28 dez. 2025.

GOOGLE. Looker Studio: visualize seus dados. [S. l.]. Disponível em: <https://lookerstudio.google.com/overview>. Acesso em: 28 dez. 2025.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). PNAD COVID19: pesquisa nacional por amostra de domicílios COVID-19. Rio de Janeiro: IBGE, 2020. Disponível em: <https://covid19.ibge.gov.br/pnad-covid/>. Acesso em: 13 out. 2025.