

IA II

Inteligência Artificial Aplicada à Saúde

IA e Dados em Saúde: explorando o SUS com vibe coding

Odontologia e cursos da área da saúde

Sumário

1. Dados: a matéria-prima da IA e da pesquisa
2. Os dados abertos do SUS: um tesouro público
3. Python e Google Colab: seu laboratório gratuito na nuvem
4. PySUS e o assistente 'PySUS No Code'
5. O fluxo de trabalho: da pergunta ao gráfico
6. Noções de análise: lendo tabelas e gráficos com senso crítico
7. Da análise ao TCC: transformando dados em pesquisa
8. Qualidade, vieses e ética no uso de dados de saúde
9. Glossário

1. Dados: a matéria-prima da IA e da pesquisa

Na disciplina IA I vimos que o aprendizado de máquina extrai regras de **dados**. Nesta disciplina, os dados assumem o papel principal — não para treinar modelos, mas para responder perguntas reais de saúde pública com apoio da IA na parte técnica. A lógica é a mesma da clínica baseada em evidências: opiniões são hipóteses; dados bem analisados são argumentos.

Dados de saúde existem em várias formas: registros administrativos (internações, procedimentos), registros vitais (nascimentos, óbitos), cadastros (estabelecimentos, profissionais) e inquéritos populacionais. Quando organizados em **tabelas** — linhas como registros, colunas como variáveis — tornam-se analisáveis por computador. Todo o nosso trabalho nesta disciplina gira em torno de obter, organizar e interpretar tabelas.

Ideia central da disciplina

Você não vai aprender a programar. Você vai aprender a fazer boas perguntas e a usar a IA para executar a parte técnica — obtendo resultados que podem, inclusive, alimentar o seu TCC.

2. Os dados abertos do SUS: um tesouro público

O Brasil possui um dos maiores acervos de dados públicos de saúde do mundo, mantido pelo **DATASUS**, o departamento de informática do SUS. São décadas de registros de todo o território nacional, de acesso gratuito, usados em milhares de artigos científicos. Conhecer os principais sistemas é essencial para escolher a fonte certa para cada pergunta:

- **SIH/SUS** — Sistema de Informações Hospitalares: internações pagas pelo SUS (diagnósticos, procedimentos, custos, tempo de permanência);
- **SIA/SUS** — Sistema de Informações Ambulatoriais: procedimentos ambulatoriais, incluindo os odontológicos;
- **SIM** — Sistema de Informações sobre Mortalidade: óbitos e suas causas (CID-10);
- **SINASC** — nascidos vivos e condições do nascimento;
- **SINAN** — agravos e doenças de notificação compulsória;
- **CNES** — cadastro de estabelecimentos e profissionais de saúde (onde estão os consultórios odontológicos e as equipes de saúde bucal do país);
- **SISAB** — informações da Atenção Primária, incluindo indicadores de saúde bucal.

Perguntas odontológicas que esses dados respondem: Quantas extrações dentárias o SUS realiza por ano em cada estado? A oferta de tratamento endodôntico na rede pública cresceu? Qual a mortalidade por câncer de boca na sua região, e como evolui no tempo? Onde faltam equipes de saúde bucal? Cada uma dessas perguntas já rendeu TCCs, dissertações e artigos.

Importante

Esses dados são públicos e anonimizados na origem (não identificam pacientes). Por isso, em geral, pesquisas com dados agregados do DATASUS dispensam aprovação de Comitê de Ética — mas confirme sempre as normas da sua instituição e da Resolução CNS 510/2016.

3. Python e Google Colab: seu laboratório gratuito na nuvem

Python é a linguagem de programação mais usada em ciência de dados e IA — justamente por ser legível e ter milhares de bibliotecas prontas. Você não precisa dominá-la: precisa apenas reconhecer o ambiente onde o código gerado pela IA será executado.

O **Google Colab** é um serviço gratuito que roda Python na nuvem, direto do navegador, sem instalar nada. O documento de trabalho é o **notebook**: uma página composta de **células** que podem conter texto explicativo ou código. Cada célula de código tem um botão de executar (play); o resultado — uma tabela, um gráfico, uma mensagem — aparece logo abaixo.

O essencial para operar o Colab

- Executar células **em ordem**, de cima para baixo (células posteriores dependem das anteriores);
- Esperar a célula terminar (o ícone giratório) antes de executar a próxima;
- Se algo travar ou sair da ordem, usar 'Ambiente de execução → Reiniciar' e executar tudo de novo — reiniciar resolve a maioria dos problemas;
- Copiar mensagens de erro e colá-las de volta na IA, que quase sempre explica e corrige.

O notebook tem uma vantagem acadêmica enorme: ele documenta o passo a passo da análise. Compartilhar o notebook é compartilhar a metodologia — qualquer colega pode reproduzir seus resultados, princípio central da ciência aberta.

4. PySUS e o assistente 'PySUS No Code'

PySUS é uma biblioteca Python criada pela comunidade científica brasileira para baixar e ler os dados do DATASUS diretamente por código — sem navegar em sites, sem converter arquivos manualmente. Ela entende os formatos técnicos usados pelo DATASUS e entrega os dados prontos em tabelas.

Para tornar esse poder acessível a quem não programa, a disciplina usa o **PySUS No Code**: um assistente construído sobre o ChatGPT que conversa em português, entende sua pergunta de pesquisa e gera o código Python **em pedaços numerados**, na ordem certa para colar e executar célula por célula no Colab. O assistente também explica o que cada pedaço faz e ajuda a interpretar os resultados e as mensagens de erro.

Por que 'em pedaços'?

Executar por partes tem função didática e prática: você vê cada etapa acontecer (baixar → filtrar → agrupar → visualizar), localiza erros com facilidade e entende a lógica da análise mesmo sem ler o código em detalhe. É o vibe coding aplicado à ciência de dados: a IA opera a técnica, você comanda o raciocínio.

Postura recomendada

Não trate o assistente como uma caixa-preta. Pergunte 'o que esta etapa faz?', 'por que esse filtro?', 'o que significa esse resultado?'. A IA é também um professor particular disponível 24 horas.

5. O fluxo de trabalho: da pergunta ao gráfico

Toda análise na disciplina segue o mesmo ciclo, que vale a pena internalizar:

- **1. Pergunta:** defina algo específico e mensurável. 'Saúde bucal no Brasil' é um tema; 'a taxa de internação por causas odontológicas no Paraná caiu entre 2015 e 2024?' é uma pergunta;
- **2. Fonte:** escolha o sistema do DATASUS que contém a resposta (capítulo 2);
- **3. Prompt:** descreva ao PySUS No Code a pergunta, o período, a região e o que deseja ver (tabela? gráfico de linha? mapa?);
- **4. Execução:** cole e rode os pedaços de código no Colab, na ordem;
- **5. Interpretação:** leia o resultado com olhar clínico e crítico — os números fazem sentido? Há saltos estranhos? O que explicaria o padrão?;
- **6. Iteração:** refine a pergunta e repita. Análises boas quase nunca saem na primeira tentativa;
- **7. Registro:** salve o notebook, anote decisões e exporte tabelas e gráficos para uso posterior (por exemplo, no TCC).

Note que os passos 1, 5 e 6 — os mais importantes — são inteiramente humanos. A IA acelera os passos técnicos, mas a pergunta, o julgamento e a conclusão são seus.

6. Noções de análise: lendo tabelas e gráficos com senso crítico

Estatística descritiva mínima

Para a maioria dos trabalhos da disciplina, basta a estatística descritiva: **contagens e frequências** (quantos procedimentos, quantos óbitos), **médias e medianas** (idade média, tempo mediano de internação), **proporções e taxas** (óbitos por 100 mil habitantes — essenciais para comparar regiões de tamanhos diferentes) e **séries temporais** (evolução ao longo dos anos).

Escolhendo o gráfico certo

- **Linha:** evolução no tempo;
- **Barras:** comparação entre categorias (estados, faixas etárias, procedimentos);
- **Pizza:** composição de um todo — use com moderação e poucas fatias;
- **Mapa temático:** distribuição geográfica de taxas.

Armadilhas clássicas de interpretação

- **Correlação não é causalidade:** duas curvas que sobem juntas não provam que uma causa a outra;
- **Números absolutos enganam:** São Paulo terá quase sempre 'mais casos' que qualquer estado — compare taxas, não totais;
- **Quedas artificiais:** registros dos meses/anos mais recentes podem estar incompletos por atraso de digitação, criando falsas quedas no fim da série;
- **Mudanças de registro:** alterações de tabelas, códigos e regras de notificação ao longo dos anos podem criar degraus que não refletem a realidade epidemiológica.

7. Da análise ao TCC: transformando dados em pesquisa

As descobertas feitas na disciplina podem — e devem — alimentar seu Trabalho de Conclusão de Curso. Estudos com dados secundários do DATASUS são um formato consolidado na literatura brasileira: são viáveis no tempo de uma graduação, não dependem de coleta em campo e tratam de problemas reais do SUS.

Estrutura típica de um estudo com dados do DATASUS

- **Introdução:** o problema de saúde e a lacuna que sua pergunta preenche;
- **Métodos:** estudo ecológico/descritivo com dados secundários; cite a fonte (sistema, ex.: SIH/SUS via DATASUS), o período, a abrangência, as variáveis e as ferramentas (Python/PySUS no Google Colab, com apoio de IA generativa);
- **Resultados:** tabelas e gráficos gerados nos notebooks, com legendas completas;
- **Discussão:** compare seus achados com a literatura científica (esta parte exige leitura de artigos — a IA pode ajudar a encontrar e resumir, mas as referências devem ser verificadas uma a uma);
- **Limitações:** declare as fragilidades dos dados administrativos (subregistro, erros de digitação, finalidade original não científica).

Transparência no uso de IA

Declare no TCC, na seção de métodos, o uso de ferramentas de IA para geração de código e apoio à redação, conforme as normas da instituição. Transparência é proteção — para você e para a ciência.

8. Qualidade, vieses e ética no uso de dados de saúde

- **Dados administrativos não nasceram para pesquisa:** registram faturamento e gestão; refletem também o acesso ao sistema, não apenas a doença. Onde não há serviço, não há registro — ausência de dado não é ausência de problema;
- **Qualidade heterogênea:** completude e precisão variam entre municípios e anos; cruzar fontes e observar tendências longas reduz o risco de conclusões frágeis;
- **Anonimização e LGPD:** trabalhe apenas com dados agregados/anonimizados; jamais tente reidentificar indivíduos, mesmo que tecnicamente possível em municípios pequenos com eventos raros;
- **Vieses de interpretação:** cuidado para não transformar diferenças de acesso em julgamentos sobre populações; contexto socioeconômico importa;
- **Alucinações da IA:** confira números e, principalmente, referências bibliográficas sugeridas por LLMs — parte delas pode não existir;
- **Responsabilidade final:** o autor do trabalho é você. A IA é ferramenta; o rigor é humano.

9. Glossário

Célula (notebook)	Bloco de um notebook que contém texto ou código executável; os resultados aparecem logo abaixo.
CID-10	Classificação Internacional de Doenças, usada para codificar diagnósticos e causas de óbito.
Dados secundários	Dados coletados originalmente para outro fim (ex.: gestão do SUS) e reutilizados em pesquisa.
DATASUS	Departamento de informática do SUS, responsável pelos sistemas nacionais de informação em saúde.
Estudo ecológico	Desenho de pesquisa cuja unidade de análise é a população/território (não o indivíduo), típico de análises com DATASUS.
Google Colab	Ambiente gratuito do Google para executar notebooks Python na nuvem, pelo navegador.
Notebook	Documento interativo que mistura texto, código e resultados; padrão da ciência de dados.
Pandas / DataFrame	Biblioteca Python para manipular tabelas; DataFrame é o nome técnico da tabela em Python.
PySUS	Biblioteca Python de código aberto para baixar e ler dados dos sistemas do DATASUS.
PySUS No Code	Assistente baseado em ChatGPT usado na disciplina, que gera código PySUS em pedaços executáveis no Colab a partir de pedidos em português.
Série temporal	Sequência de valores ao longo do tempo (ex.: internações por ano).
Taxa	Medida relativa (ex.: casos por 100 mil habitantes) que permite comparar populações de tamanhos diferentes.
TabNet	Ferramenta oficial de tabulação online do DATASUS; alternativa sem código para consultas simples.