



Instituto Infnet

GRADUAÇÃO EM BANCO DE DADOS

Projeto de Bloco: Ciência de Dados Aplicada [24E3_5]

WANDERSON RAFAEL MENDONÇA BATISTA

TESTE DE PERFORMANCE – TP5

PROF. DIEGO DA SILVA RODRIGUES

RIO DE JANEIRO, 2024

Configuração do Ambiente de Desenvolvimento:

Configure seu ambiente de desenvolvimento, incluindo Git para controle de versão e preparação para deploy. Lembre-se de seguir a estrutura do CRISP-DM para organizar seu projeto de forma eficiente e escalável.

<https://github.com/wanderaf/ProjetoDeBloco>

1. Atualização e Completação dos Artefatos do TSDP (Team Data Science Process):

- **Revise e complete todos os artefatos do TDSP, que orientaram seu processo de ciência de dados ao longo do projeto.**
- **Data Summary Report: Detalhe como os dados coletados via APIs ou web scraping foram integrados e utilizados ao longo do projeto.**
- **Project Charter: Atualize e argumente sobre como os dados, o uso de IA e a engenharia de prompts estão alinhados com a resolução do problema de negócio escolhido, destacando sua relevância no contexto de Ciência de Dados.**
- **Certifique-se de que a documentação reflita o ciclo completo de ciência de dados, desde a obtenção dos dados até a transformação deles em insights por meio de IA**

Os arquivos foram atualizados e disponibilizados no git

2. Escolha de um Caminho para a Solução Final com LLMs (Local ou via API): O aluno deve optar por um dos três caminhos abaixo, integrando-o de maneira eficaz ao ciclo de ciência de dados do projeto. A solução deve fazer uso dos dados coletados, aplicando as técnicas de IA e engenharia de prompts no processamento e análise dessas informações.

- **Opção 1: Implementação de Memória Conversacional com LLMs**
Integre LLMs para implementar memória conversacional na sua aplicação, permitindo que o sistema mantenha o contexto ao longo de várias interações.
Ciência de Dados Aplicada: A memória conversacional pode ser aplicada para interpretar perguntas e manter o histórico de interações ao lidar com grandes volumes de dados sobre sustentabilidade ou

governança, fornecendo insights contínuos e contextualizados a partir dos dados analisados.

Exemplo: O sistema pode relembrar informações anteriores e adaptar suas respostas com base nos dados coletados sobre ESG, com a memória sendo gerenciada por modelos locais ou APIs de IA online como o ChatGPT.

- **Opção 2: Automação de Sumarização de Textos com LLMs**
Utilize LLMs para automatizar a sumarização de grandes volumes de dados textuais (como relatórios ou artigos sobre ESG).
Ciência de Dados Aplicada: Aqui, o foco é em sintetizar informações complexas e densas, transformando dados coletados em resumos claros e acionáveis, facilitando a tomada de decisões baseada em dados.

Exemplo: Ao usar dados textuais obtidos via web scraping ou APIs, como relatórios de sustentabilidade, o sistema pode gerar resumos automáticos, tornando o processamento e a análise de grandes documentos mais eficiente. A sumarização pode ser realizada com modelos locais ou APIs de IA online como OpenAI.

- **Opção 3: Desenvolvimento de Agentes Inteligentes para Tomada de Decisão**

Implemente agentes inteligentes capazes de resolver problemas complexos com base nos dados coletados.

Ciência de Dados Aplicada: Esses agentes podem automatizar a análise dos dados coletados sobre indicadores de desempenho ESG, sugerindo recomendações ou previsões baseadas em padrões encontrados nos dados.

Exemplo: Um agente pode ser configurado para analisar dados financeiros e de sustentabilidade, propondo ações corretivas ou recomendações baseadas em benchmarks de governança, utilizando modelos locais ou APIs de IA online.

O projeto se alinha à Opção 3: Desenvolvimento de Agentes Inteligentes para Tomada de Decisão, utilizando o modelo t5-small localmente para analisar dados de saúde armazenados em MongoDB e realizar previsões sobre valores e quantidades aprovadas

em procedimentos médicos. A integração de técnicas de engenharia de prompts com LLMs permite identificar padrões nos dados e fornecer insights preditivos e visuais para suporte à tomada de decisão.

3. **Desenvolvimento de um Dashboard Final com Modelos de IA:** Após implementar a funcionalidade escolhida (memória conversacional, sumarização ou agente inteligente), integre-a a um **dashboard interativo** que demonstre claramente o ciclo de **Ciência de Dados**, desde a coleta dos dados até a geração de insights. O dashboard deve:
 - Exibir os **dados coletados e processados** ao longo do projeto, mostrando como os modelos de IA foram aplicados para transformar esses dados em conhecimento útil.
 - Oferecer **visualizações e gráficos** que ilustrem os principais resultados, insights ou recomendações gerados pelos modelos de IA.
 - Permitir que o usuário interaja com os dados e modelos de IA, seja utilizando **modelos locais** ou **APIs de IA online**, para obter respostas em tempo real baseadas no contexto de **sustentabilidade e governança**.

Abaixo algumas imagens do dashboard

Visualização de Dados

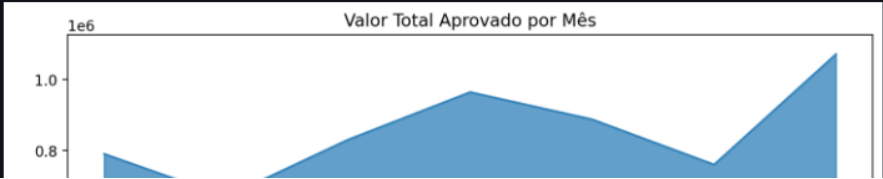
Carregar Dados Filtrados

Dados carregados com sucesso!

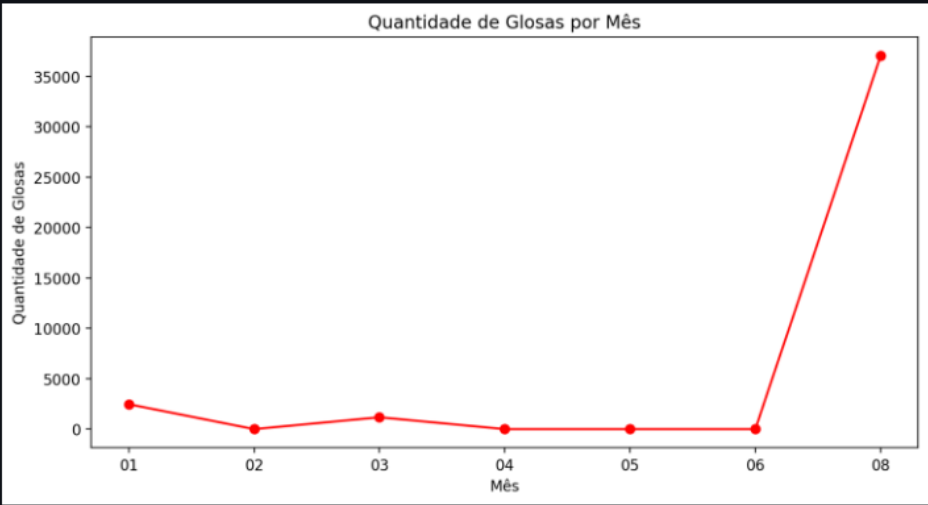
Resumo Estatístico:

	TOTAL_PA_QTDPRO	TOTAL_PA_QTDAPR	TOTAL_PA_VALPRO	TOTAL_PA_VALAPR	GLOSAS
count	1,473	1,473	1,473	1,473	1,473
mean	242.3781	239.7984	4,083.5365	4,055.9205	27.616
std	499.6858	492.6714	9,860.6565	9,824.1729	798.5393
min	1	0	1.53	0	0
25%	12	12	136.22	135	0
50%	63	62	806.55	806.25	0
75%	246	245	3,268.95	3,240	0
max	4,891	4,891	111,099.6	111,099.6	30,000

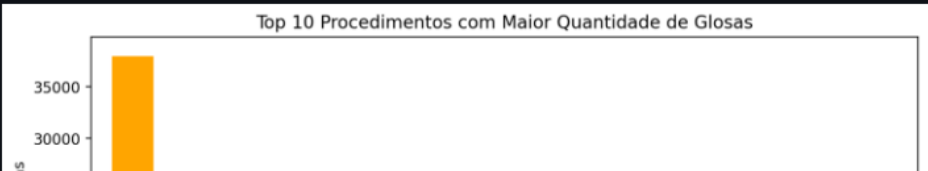
Evolução do Valor Aprovado por Mês



Quantidade de Glosas por Mês



Principais Procedimentos com Glosas



6	Alto Alegre do Maranhão	210043
7	Alto Alegre do Pindaré	210047
8	Alto Parnaíba	210050
9	Amapá do Maranhão	210055

Selecionar Município

Açailândia

Município selecionado: Açailândia (Código: 210005)

Selecionar Procedimento

0408020199 - REDUCAO INCRUENTA DE FRATURA DA DIAFISE DO UMEMO

Buscar Unidades que Realizam o Procedimento

Unidades encontradas!

Número de unidades encontradas: 1

	PA_CODUNI	FANTASIA	TOTAL_PA_VALPRO
0	2463016	HOSPITAL MUNICIPAL DE ACAILANDIA	75.7600

API - Mongo: Consulta e Inserção de Unidades de Saúde

Consulta (GET) Inserção (POST)

Inserção de Unidades

Código da Unidade (PA_CODUNI):

Nome Fantasia:

Total da Quantidade Produzida (TOTAL_PA_QTDPRO):

 - +

Total da Quantidade Aprovada (TOTAL_PA_QTDAPR):

 - +

Valor Produzido (TOTAL_PA_VALPRO):

 - +

Valor Aprovado (TOTAL_PA_VALAPR):

4. Apresentação e Argumentação da Solução Final:

- Além da implementação técnica, você deverá **apresentar a solução** e **justificar** como ela resolve o problema de negócio, destacando a importância de cada etapa do ciclo de ciência de dados (da coleta e limpeza à modelagem e apresentação).
- Destaque o uso de **dados reais** e como os modelos de IA aplicados oferecem insights relevantes para a governança e sustentabilidade, alinhando a solução aos **ODS** e práticas de **ESG**.
- A simplicidade no uso do dashboard, combinada com a profundidade dos insights gerados, será um critério importante na avaliação.

A solução desenvolvida é um dashboard interativo implementado em Streamlit, baseado em dados reais do SUS coletados do estado do Maranhão e armazenados em MongoDB. O projeto segue todas as etapas do ciclo de ciência de dados, incluindo coleta, limpeza, modelagem com o modelo t5-small e apresentação visual (desde a coleta de dados no ftp

do SUS, tratamento e inclusão de informações no MongoDB, conexão entre banco e streamlit, bem como utilização de inteligência artificial para prever quantitativo e valores de procedimentos das unidades)

Os gráficos e tabelas permitem análises como evolução de valores aprovados, identificação de glosas e comparação entre valores produzidos e aprovados, fornecendo insights práticos para melhorar a alocação de recursos na saúde pública.

Alinhado aos ODS 3 (Saúde e Bem-Estar) e 16 (Paz, Justiça e Instituições Eficazes), o projeto promove eficiência, transparência e sustentabilidade na gestão da saúde, ajudando o Maranhão a superar desafios e otimizar o uso de seus recursos limitados.