
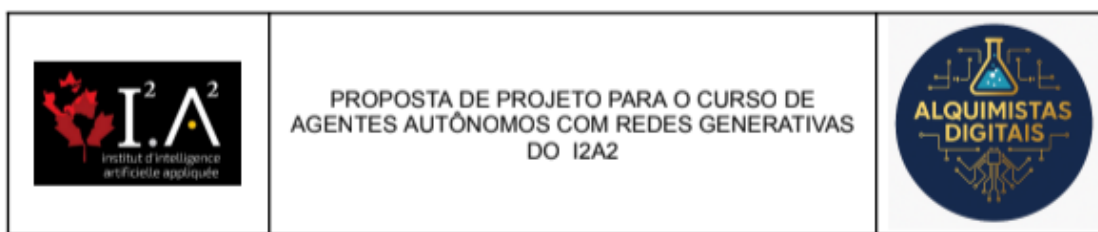
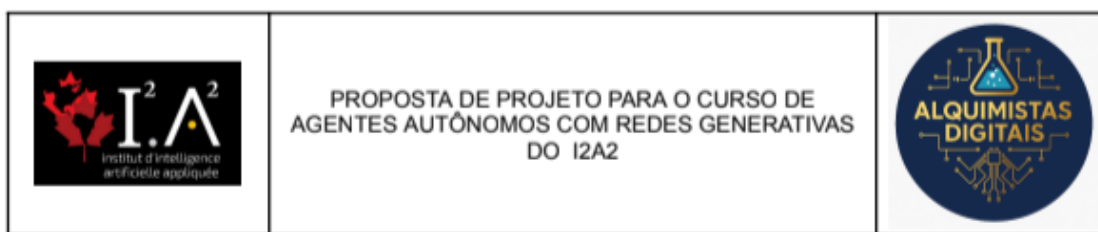
	<p>PROPOSTA DE PROJETO PARA O CURSO DE AGENTES AUTÔNOMOS COM REDES GENERATIVAS DO I2A2</p>	
---	--	---

Relatório da Atividade Extra:
EDAna – Agente E.D.A (Exploração de Dados
Automatizada)



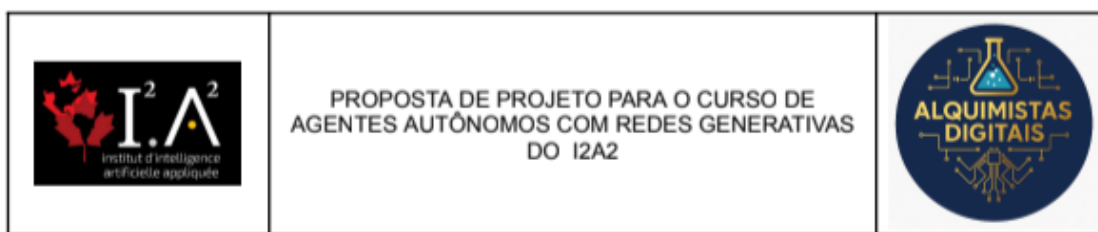
1. Identificação do projeto discente

Título do projeto: Agentes Autônomos - Relatório da Atividade Extra	
Professor responsável: Celso Azevedo	
Curso: Agentes Autônomos com Redes Generativas	
Atividade: Setembro/2025	
Nome do grupo: Alquimistas Digitais	
Integrante	E-mail
Amanda Praça	amandapraça@gmail.com



2. Sumário

1.Introdução
2.Framework escolhida
3.Estrutura da Solução
4.Componentes Principais
5.Funcionamento do Agente
6.Interface do Agente Autônomo – EDAna
7. Visualizações gráficas disponíveis
8.Funcionalidades da Interface
9.Perguntas realizadas e respostas do agente
10.Conclusões do Agente
11.Códigos fontes utilizados
12.Link de acesso ao agente
13.Considerações Finais
14.Referências



DADOS DO PROJETO DISCENTE

1. INTRODUÇÃO

Este relatório apresenta o desenvolvimento do agente autônomo **EDAna**, uma especialista em análise exploratória de dados.

O projeto foi realizado de forma individual pela aluna Amanda Praça, como parte do desafio proposto na disciplina de Agentes Autônomos com Redes Generativas.

A **EDAna** foi desenvolvida com o objetivo de permitir que usuários façam upload de arquivos CSV e obtenham análises automatizadas, respostas inteligentes e visualizações gráficas com base nos dados fornecidos.

2. FRAMEWORK ESCOLHIDA

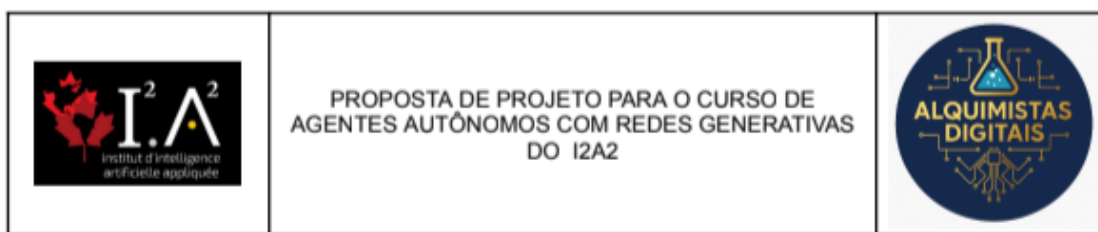
A solução foi construída utilizando as seguintes tecnologias:

- **Python 3.10+** como linguagem principal;
- **Streamlit** para construção da interface web interativa;
- **Gemini (via API)** como modelo de linguagem para geração de respostas;
- **Pandas** para manipulação de dados;
- **Matplotlib e Seaborn** para geração de gráficos;
- **dotenv** para gerenciamento seguro da chave da API.

3. COMPONENTES PRINCIPAIS

A seguir, são apresentados os principais componentes que estruturam o funcionamento do agente **EDAna**:

- **Interface Streamlit:** Permite upload de arquivos CSV e entrada de perguntas pelo usuário em uma interface visual amigável e responsiva;
- **Análise automatizada com pandas, seaborn e matplotlib:** O agente gera estatísticas descritivas, gráficos interativos e resumos dos dados de forma autônoma;
- **Integração com LLM Gemini** A LLM é acionada automaticamente para interpretar os dados e responder perguntas com base no conteúdo do CSV;
- **Prompt estruturado:** O agente envia à LLM um resumo técnico dos dados, garantindo que as respostas sejam baseadas exclusivamente nas informações disponíveis;
- **Memória local:** O agente registra perguntas e respostas anteriores, permitindo que construa conclusões e evite chamadas repetidas à API;



- **Segurança:** A chave da API do Gemini é armazenada em arquivo .env, oculta do código-fonte e do repositório público.

4. FUNCIONAMENTO DO AGENTE

O funcionamento da **EDAna** segue um fluxo autônomo e generativo, conforme os passos abaixo:

Passo a passo:

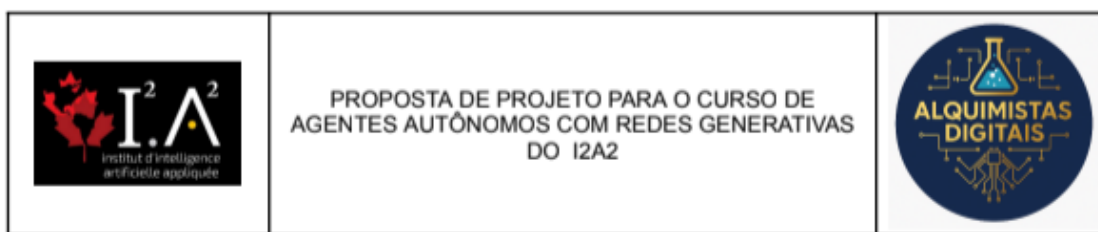
1. O usuário acessa o agente via link público;
2. Faz upload de um arquivo CSV ou seleciona um já disponível;
3. Digita uma pergunta livre sobre os dados;
4. O agente processa o CSV, gera um resumo técnico e envia à LLM;
5. A resposta é exibida na interface, podendo incluir gráficos gerados com código Python;
6. As análises são registradas em memória, permitindo que o agente construa conclusões ao longo da interação.

Essa abordagem garante que o agente seja autônomo, generativo e confiável, respondendo com base nos dados e sem intervenção manual. A LLM está integrada ao código e opera como parte do agente, conforme exigido pela atividade.

5. ESTRUTURA DA SOLUÇÃO

A solução foi estruturada em três arquivos principais:

- **app.py:** é o núcleo da aplicação Streamlit. Ele gerencia a interface do usuário, o upload de arquivos CSV, a visualização dos dados, a geração de gráficos e a interação com o agente EDAna. Também organiza as abas de navegação (“Análise Interativa” e “Conclusões da EDAna”) e controla o histórico de perguntas e respostas;
- **llm_utils.py:** contém as funções responsáveis por gerar respostas usando o modelo Gemini. Ele extrai estatísticas dos dados, monta o prompt com instruções específicas e envia a pergunta para a LLM, retornando uma resposta contextualizada com base no conteúdo do CSV;
- **agente.eda.py:** reúne funções auxiliares que complementam a análise exploratória. Entre elas, destaca-se a função `gerar_conclusoes()`, que interpreta o histórico de perguntas feitas pelo usuário e gera insights automáticos com base em padrões detectados (como distribuição, outliers, correlação e recorrência de valores).

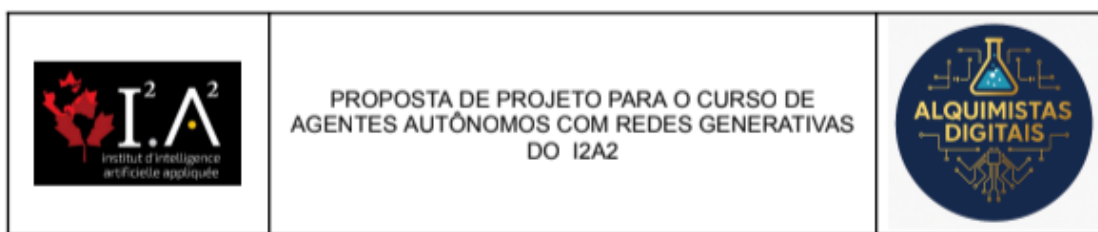


Essa estrutura modular permite que a aplicação seja organizada, escalável e facilmente adaptável para novos tipos de análise ou visualização. A lógica da aplicação permite que o usuário carregue um arquivo CSV, visualize os dados, faça perguntas sobre o conteúdo e gere gráficos com base nas colunas numéricas. As respostas são adaptadas conforme o estilo desejado: por padrão, são técnicas e completas, mas o usuário pode solicitar uma **(resposta curta)** ao final da pergunta para obter uma versão mais objetiva.

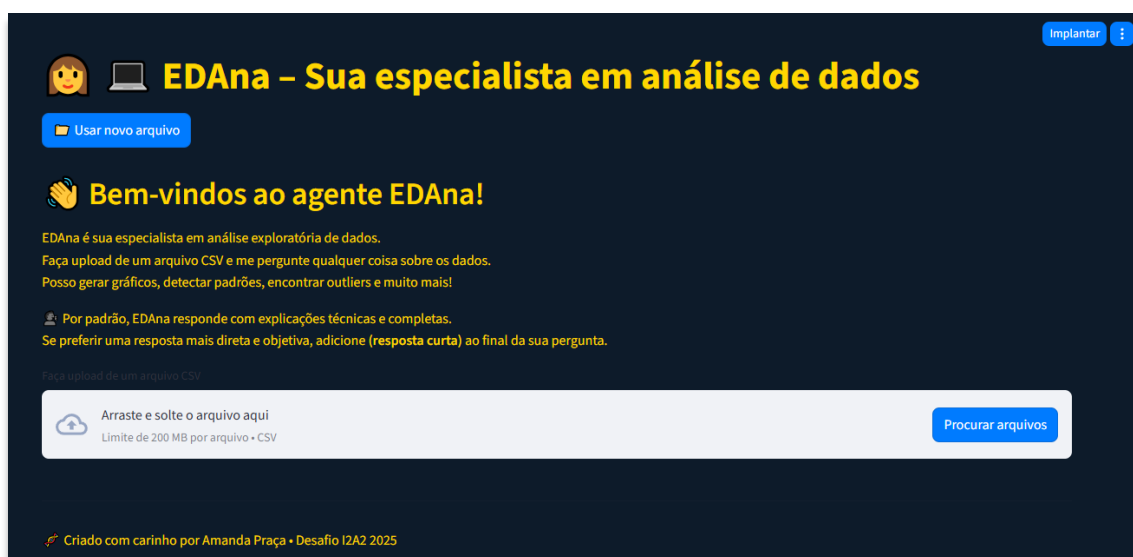
6. INTERFACE DO AGENTE AUTÔNOMO

A interface da **EDAna** foi desenvolvida com foco em acessibilidade, sofisticação e interação natural. Utilizando a biblioteca **Streamlit**, o agente apresenta uma estrutura visual intuitiva, permitindo que o usuário explore os dados sem necessidade de comandos técnicos.

- **Identidade visual:** A paleta de cores foi personalizada para refletir uma identidade visual própria, com **fundo escuro** e **elementos em dourado**, reforçando a ideia de sofisticação, inteligência e exclusividade. Ícones e mensagens de boas-vindas foram adicionados para tornar a experiência mais acolhedora;
- **Entrada em linguagem natural:** O agente permite que o usuário digite perguntas livres sobre os dados, utilizando linguagem natural. As respostas são geradas de forma contextualizada, com base no conteúdo do arquivo CSV, sem necessidade de comandos específicos ou termos técnicos;



- **Campos da interface:**



📁 Seção: Gerenciamento de Arquivos CSV

A interface da **EDAna** oferece uma experiência intuitiva para o carregamento e substituição de arquivos CSV, permitindo que o usuário explore diferentes conjuntos de dados de forma dinâmica e sem complicações.

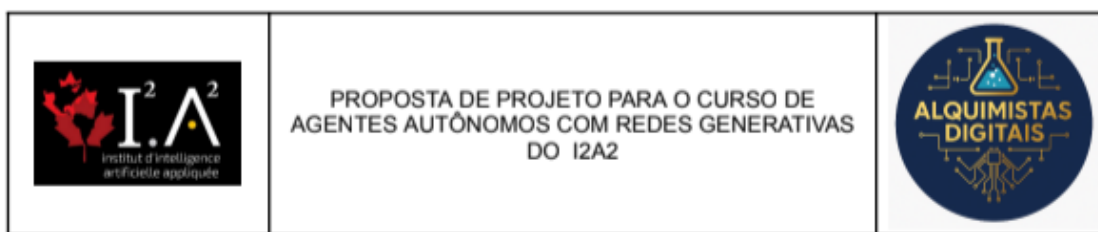
🔄 Botão “Usar novo arquivo”

Este botão permite ao usuário **reiniciar a sessão de análise**, limpando o histórico de perguntas e removendo o arquivo CSV atual. Ao clicar, a aplicação é recarregada e retorna ao estado inicial, pronta para receber um novo conjunto de dados.

Objetivo:

- Facilitar a troca de arquivos sem necessidade de reiniciar manualmente o aplicativo;
- Garantir que análises anteriores não interfiram na nova exploração.

📁 Área de Upload de Arquivo



Logo abaixo da introdução, o usuário encontra a área de upload com suporte a arquivos CSV de até **200 MB**. A interface oferece duas formas de envio:

- **Arrastar e soltar** diretamente na área destacada;
- **Botão “Procurar arquivos”**, que abre o explorador de arquivos do sistema.

Assim que o arquivo é carregado, a **EDAna** realiza uma leitura inteligente dos dados, identificando colunas, tipos de variáveis e corrigindo automaticamente formatações incorretas (como separadores mal definidos ou cabeçalhos duplicados).

Destaques técnicos:

- Reconhecimento automático de colunas numéricas.
- Correção de arquivos mal formatados.
- Feedback visual com mensagens de sucesso ou erro.

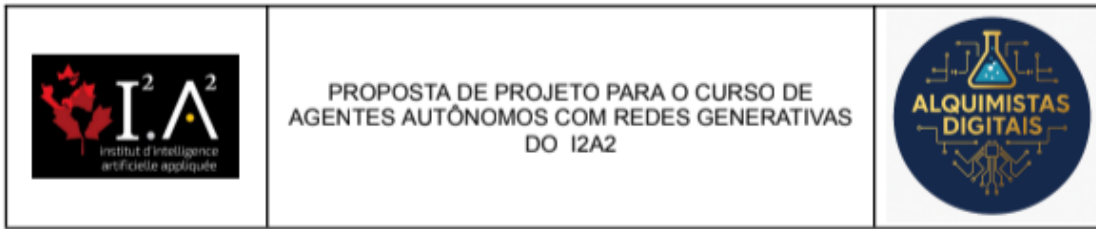
✓ Benefícios para o usuário

- **Flexibilidade** para testar diferentes bases de dados;
- **Segurança** ao limpar o histórico anterior com um clique;
- **Acessibilidade** com suporte a arrastar/soltar e seleção manual.

Pré-visualização dos dados

	Tipo	Bairro	Quartos	Vagas	Suites	Area	Valor	Condominio	IPTU
0	Quitinete	Copacabana	1	0	0	40	1700	500	60
1	Casa	Jardim Botânico	2	0	1	100	7000	None	None
2	Conjunto Comercial/Sala	Barra da Tijuca	0	4	0	150	5200	4020	1111
3	Apartamento	Centro	1	0	0	15	800	390	20
4	Apartamento	Higienópolis	1	0	0	48	800	230	None

🕒 Seção: Pré-visualização e Navegação Interativa



A interface da **EDAna** oferece uma navegação simples e eficiente por meio de abas que organizam as funcionalidades principais da ferramenta:

◆ Abas de Navegação

- **Análise Interativa:** Esta aba concentra os recursos de exploração ativa dos dados. O usuário pode visualizar o conteúdo do CSV, fazer perguntas diretamente à **EDAna** e gerar gráficos com base nas variáveis disponíveis.
- **Conclusões da EDAna:** Nesta aba, são exibidos os insights automáticos gerados pela **EDAna** com base nas perguntas feitas anteriormente. A ferramenta sintetiza padrões, correlações e comportamentos detectados nos dados, oferecendo uma visão resumida e estratégica.

👁 Pré-visualização dos Dados

Logo abaixo das abas, a interface exibe uma tabela com os dados carregados do arquivo CSV. Essa pré-visualização permite ao usuário:

- Verificar se o arquivo foi lido corretamente;
- Identificar os nomes das colunas e os tipos de variáveis disponíveis;
- Observar os primeiros registros para entender a estrutura do dataset.

No exemplo apresentado, os dados referem-se a imóveis para aluguel, com colunas como:

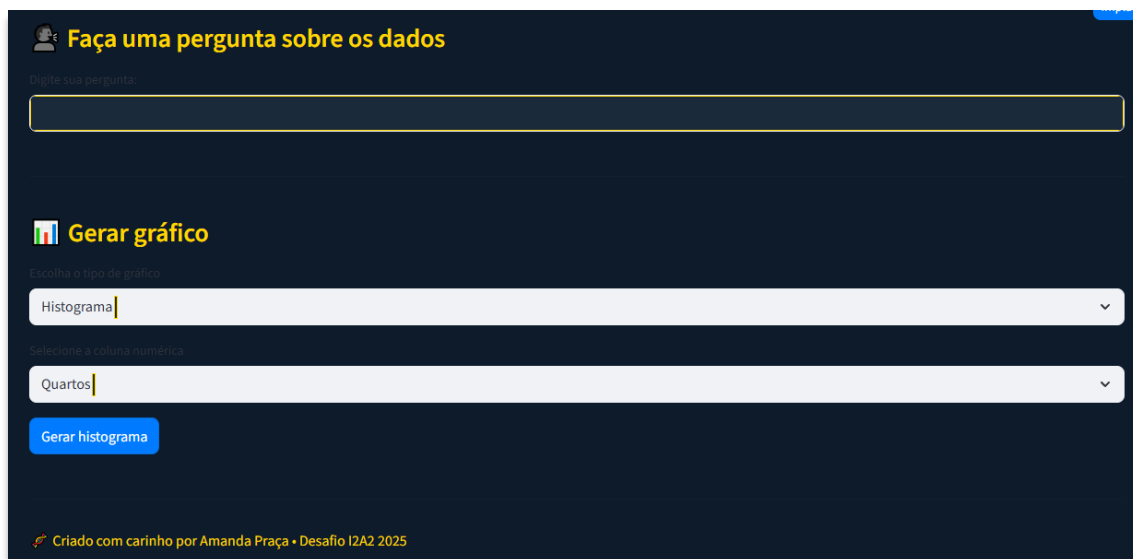
- **Tipo:** tipo de imóvel (ex: apartamento, casa, quitinete).
- **Bairro:** localização do imóvel.
- **Quartos, Vagas, Suítes:** características físicas.
- **Área:** metragem do imóvel.
- **Valor, Condomínio, IPTU:** valores financeiros associados.

Essa visualização inicial é essencial para orientar o usuário na formulação de perguntas e na escolha de variáveis para os gráficos.

✓ Benefícios para o usuário


- **Confirmação visual** de que o arquivo foi carregado corretamente;
- **Facilidade para explorar** os dados sem precisar abrir o CSV externamente;

- **Base para decisões analíticas**, como escolher colunas relevantes para perguntas ou gráficos.



Seção: Geração de Gráficos Interativos

A funcionalidade "**Gerar gráfico**" permite ao usuário visualizar rapidamente padrões, distribuições e relações entre variáveis numéricas do conjunto de dados carregado. Assim que o arquivo CSV é importado, EDAna identifica automaticamente as colunas com dados numéricos e as disponibiliza para seleção.

 Como funciona:

- O usuário escolhe o tipo de gráfico desejado entre:
 - **Histograma**: mostra a distribuição de valores em uma única coluna.
 - **Boxplot**: destaca a mediana, quartis e possíveis outliers.
 - **Dispersão (Scatterplot)**: revela correlações ou padrões entre duas variáveis numéricas.
- Em seguida, o usuário seleciona a(s) coluna(s) que deseja visualizar.
 - Para histogramas e boxplots: uma única coluna numérica.
 - Para gráficos de dispersão: duas colunas numéricas (eixo X e eixo Y).
- Ao clicar no botão correspondente (ex: "Gerar histograma"), o gráfico é renderizado dinamicamente com base nos dados reais do CSV.

✓ Benefícios:

- **Exploração visual rápida:** sem necessidade de escrever código.
- **Adaptação automática:** os nomes das colunas são extraídos diretamente do CSV carregado.
- **Facilidade de uso:** ideal para usuários não técnicos que desejam entender os dados de forma visual.

7. VISUALIZAÇÕES GRÁFICAS DISPONÍVEIS

O agente EDAna é capaz de gerar diferentes tipos de gráficos com base nas variáveis numéricas do arquivo CSV. Esses gráficos ajudam a visualizar padrões, detectar anomalias e compreender relações entre variáveis.

Histograma

- **O que faz:** mostra a distribuição de frequência dos valores de uma variável.
- **Para que serve:** identificar concentração de dados, assimetrias e possíveis outliers.
- **Exemplo de uso:** verificar se a maioria das transações está abaixo de determinado valor.

Boxplot

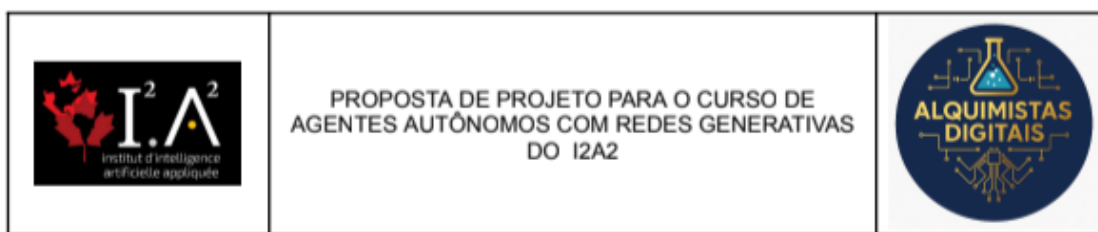
- **O que faz:** exibe a mediana, quartis e possíveis outliers de uma variável.
- **Para que serve:** detectar valores atípicos e entender a dispersão dos dados.
- **Exemplo de uso:** avaliar se há transações com valores muito acima da média.

Gráfico de dispersão

- **O que faz:** plota duas variáveis numéricas em um plano cartesiano.
- **Para que serve:** identificar correlações, agrupamentos e padrões entre variáveis.
- **Exemplo de uso:** verificar se há relação entre o valor da transação e o tempo.

8. FUNCIONALIDADE DA INTERFACE

A interface da **EDAna** foi projetada para ser intuitiva e acessível.



As principais seções são:

- **Upload de arquivo CSV:** permite carregar arquivos de até 200 MB, com suporte a arrastar e soltar ou seleção manual;
- **Botão “Usar novo arquivo”:** reinicia a sessão, limpando o histórico e permitindo nova análise;
- **Análise Interativa:** exibe os dados, permite perguntas e gera gráficos;
- **Conclusões da EDAna:** apresenta insights automáticos com base nas perguntas feitas;
- **Geração de gráficos:** permite criar histogramas, boxplots e gráficos de dispersão com base nas colunas numéricas do CSV.

9. PERGUNTAS REALIZADAS E RESPOSTAS

Nesta seção, são apresentadas as perguntas feitas ao agente **EDAna** durante o processo de análise exploratória dos dados. As respostas foram geradas automaticamente com base no conteúdo do arquivo CSV, utilizando o modelo de linguagem Gemini integrado ao sistema.

Por padrão, o agente responde com explicações **técnicas e completas**, oferecendo interpretações detalhadas dos dados e dos padrões identificados. No entanto, caso o solicitante prefira uma resposta mais direta e objetiva, basta adicionar o termo “**(resposta curta)**” ao final da pergunta. Essa funcionalidade permite adaptar o estilo da resposta conforme a necessidade do usuário.

A seguir, são apresentados exemplos de perguntas realizadas, incluindo uma comparação entre os dois estilos de resposta (**Pergunta 1**):

Pergunta 1: “Existe correlação entre as variáveis?”

Resposta padrão:



PROPOSTA DE PROJETO PARA O CURSO DE
AGENTES AUTÔNOMOS COM REDES GENERATIVAS
DO I2A2



1. Pergunta: Existe correlação entre as variáveis?

Resposta: Sim, existe correlação entre as variáveis, mas a intensidade e a direção variam consideravelmente dependendo do par de variáveis analisado.

Com base nos dados fornecidos, podemos observar os seguintes pontos:

- Correlação entre as variáveis "V":** As variáveis de "V1" a "V28" apresentam correlação praticamente nula entre si. Os coeficientes de correlação entre elas são extremamente próximos de zero (ex: a correlação entre "V1" e "V2" é de). Este padrão é típico de dados que passaram por uma transformação como a Análise de Componentes Principais (PCA), onde os componentes resultantes são ortogonais (não correlacionados). `4.13e-16`
- Correlação com a variável "Amount":** A variável "Amount" possui correlações de intensidade moderada com algumas das variáveis "V".
 - Correlação Negativa:** A correlação mais forte é com "V2" (-0.53), seguida por "V5" (-0.39) e "V1" (-0.23). Isso indica que, em geral, um aumento nos valores dessas variáveis está associado a uma diminuição no valor de "Amount".
 - Correlação Positiva:** As correlações positivas mais relevantes são com "V7" (0.40), "V20" (0.34) e "V6" (0.22). Nestes casos, um aumento no valor da variável "V" tende a estar associado a um aumento no valor de "Amount".
- Correlação com a variável "Time":** A variável "Time" também apresenta algumas correlações, sendo a mais notável com "V3" (-0.42). Isso sugere uma relação negativa moderada entre o tempo e o componente "V3".
- Variável "Class":** Os dados não permitem analisar a correlação com a variável "Class", pois suas estatísticas e coeficientes de correlação são apresentados como (Not a Number), indicando ausência de dados válidos para o cálculo. `NaN`

Resposta curta e objetiva:

2. Pergunta: Existe correlação entre as variáveis? (resposta curta)

Resposta: Sim, existe correlação entre algumas variáveis.

A correlação negativa mais forte é entre "V2" e "Amount" (-0.53). A variável "Time" também possui uma correlação negativa considerável com "V3" (-0.42).

Existem também correlações positivas, como entre "Amount" e "V7" (0.40).



PROPOSTA DE PROJETO PARA O CURSO DE AGENTES AUTÔNOMOS COM REDES GENERATIVAS DO I2A2



Pergunta 2: “Quais são as medidas de tendência central (média, mediana)?”

2. Pergunta: Quais são as medidas de tendência central (média, mediana)?

Resposta: Com base nas estatísticas descritivas fornecidas, é possível extrair as medidas de tendência central (média e mediana) para as variáveis numéricas.

Média

A média está disponível na linha da tabela de estatísticas. Ela representa o valor médio de cada variável. `mean`

- **Time:** A média é de aproximadamente 94813.86.
- **"Amount":** O valor médio das transações é 88.35.
- **Variáveis "V":** As médias das variáveis de "V1" a "V28" são todas extremamente próximas de zero (ex: "V1" tem média de 1.17e-15). Isso sugere que esses dados passaram por um processo de padronização, onde a média foi subtraída de cada valor.

Mediana (percentil 50°)

A mediana corresponde ao 50° percentil, que divide a distribuição dos dados ao meio. Ela está disponível na linha da tabela. `50%`

- **Time:** A mediana é 84692.0.
- **"Amount":** O valor mediano das transações é 22.00.
- **"V1":** A mediana é 0.018.



PROPOSTA DE PROJETO PARA O CURSO DE AGENTES AUTÔNOMOS COM REDES GENERATIVAS DO I2A2



Pergunta 3: “Qual a variabilidade dos dados (desvio padrão, variância)”?

3. Pergunta: Qual a variabilidade dos dados (desvio padrão, variância)?

Imple

Resposta: Com base nos dados fornecidos, a variabilidade das variáveis pode ser analisada através do desvio padrão (disponível na tabela de estatísticas descritivas) e da variância (que pode ser calculada como o quadrado do desvio padrão).

Desvio Padrão (std)

O desvio padrão mede a dispersão dos dados em relação à média. Um valor alto indica que os dados estão mais espalhados.

- **Alta Variabilidade:** As variáveis e apresentam a maior variabilidade. `Time`, `"Amount"`
 - `Time`: Possui um desvio padrão de **47488.15**, indicando que as transações estão amplamente distribuídas ao longo do período de tempo registrado.
 - `"Amount"`: Tem um desvio padrão de **250.12**. Considerando que a média é 88.35, este valor é bastante elevado, sugerindo que, embora a maioria das transações seja de baixo valor (75% são abaixo de 77.16), existem transações com valores muito altos que aumentam a dispersão.
- **Variabilidade Controlada:** As variáveis a , que provavelmente são resultado de uma Análise de Componentes Principais (PCA), exibem uma variabilidade muito menor e mais uniforme. `V1`, `V28`
 - A maioria dessas variáveis tem um desvio padrão próximo de **1.0** (ex: com 1.09, com 0.99). `"V9"`, `"V12"`
 - A variável com menor desvio padrão é a (**0.33**), e a com maior desvio é a (**1.95**). Isso sugere que os dados foram escalonados (padronizados) durante o pré-processamento. `"V28"`, `"V1"`

Variância

A variância é o quadrado do desvio padrão e quantifica a dispersão na unidade original ao quadrado.

- `"Amount"`: A variância é aproximadamente **62.560** (250.12^2), o que reforça a alta dispersão dos valores das transações.
- `"V1"`: A variância é de **3.84** (1.96^2).
- `"V28"`: A variância é de **0.11** (0.33^2), confirmando sua baixa dispersão em comparação com as outras variáveis.

Em resumo, os dados mostram uma grande variabilidade nas variáveis originais e , enquanto os componentes a apresentam uma dispersão significativamente menor e mais controlada. A variável não possui estatísticas de variabilidade, pois todos os seus valores são (Not a Number) na tabela descritiva fornecida. `Time`, `Amount`, `V1`, `V28`, `Class`, `NaN`

Pergunta 4: “Existem padrões ou tendências temporais”?

4. Pergunta: Existem padrões ou tendências temporais?

Resposta: Com base nos dados fornecidos, é possível identificar alguns padrões e tendências temporais, embora a maioria seja fraca. A análise se baseia na correlação entre a variável e as demais. `Time`

- Tendência Mais Significativa:** A variável apresenta a correlação negativa mais forte com `V3`, com um coeficiente de `-0.419`. Isso indica uma tendência moderada de que, à medida que o tempo passa, os valores de `V3` tendem a diminuir. `Time` `V3`
- Tendências Fracas:** Outras variáveis também mostram correlações, mas de forma mais fraca.
 - Correlações Negativas Fracas:** As variáveis `V11` e `V25` demonstram uma leve tendência a diminuir com o tempo.
 - Correlações Positivas Fracas:** As variáveis `V5` e `V22` mostram uma leve tendência a aumentar com o tempo.
- Ausência de Tendência Temporal Relevante:**
 - A maioria das outras variáveis (`V1`, `V2`, `V4`, etc.) possui coeficientes de correlação muito próximos de zero, indicando ausência de uma tendência linear temporal significativa.
 - A variável (valor da transação) tem uma correlação com `Time` de `-0.010`, que é insignificante. Isso significa que o valor das transações não tende a aumentar ou diminuir ao longo do tempo analisado.
- Impossibilidade de Análise:**
 - Não é possível analisar a tendência temporal para a variável `Class`, pois seu valor de correlação é `NaN`. Isso sugere que todos os registros no conjunto de dados pertencem à mesma classe, impedindo o cálculo de uma correlação.

Em resumo, existe um padrão temporal moderado para a variável `V3` e padrões fracos para algumas outras. No entanto, para a maioria das variáveis, incluindo o valor da transação (`Amount`), não há uma tendência temporal linear clara.

Pergunta 5: “Quais os valores mais frequentes ou menos frequentes”?

5. Pergunta: Quais os valores mais frequentes ou menos frequentes?

Resposta: Com base nos dados fornecidos, a resposta é:

Os dados não permitem responder essa pergunta.

Justificativa Técnica:

Os dados apresentados são resumos estatísticos (média, desvio padrão, quartis, etc.) e uma matriz de correlação. Para determinar os valores mais ou menos frequentes (a moda ou a distribuição de frequência), seria necessário ter acesso aos dados brutos de cada transação, o que não foi fornecido.

No entanto, podemos inferir sobre a distribuição dos valores a partir das estatísticas descritivas, especialmente para a coluna `Amount`:

- Concentração de Valores Baixos:** A mediana (quartil 50%) do valor da transação (`Amount`) é `22.00`, enquanto 75% das transações têm valor de até `77.16`. Isso indica que a grande maioria das transações envolve valores baixos.
- Presença de Valores Altos e Incomuns:** A média do valor da transação é `88.35`, que é significativamente maior que a mediana. Isso, combinado com o valor máximo de `25,691.16`, sugere que existem transações com valores muito altos que são, provavelmente, as menos frequentes (outliers) e que "puxam" a média para cima.
- Valor Mínimo:** O valor mínimo é `0.00`. Embora seja um extremo, não é possível saber a sua frequência.

Em resumo, não é possível identificar o valor específico mais frequente, mas os dados indicam que valores baixos (abaixo de `77.16`) são muito comuns, enquanto valores muito altos (próximos do máximo) são provavelmente raros.



PROPOSTA DE PROJETO PARA O CURSO DE AGENTES AUTÔNOMOS COM REDES GENERATIVAS DO I2A2



Pergunta 6: “Existem agrupamentos (clusters) nos dados”?

6. Pergunta: Existem agrupamentos (clusters) nos dados?

Impl

Resposta: Com base nos dados fornecidos, não é possível confirmar ou negar a existência de agrupamentos (clusters) de forma conclusiva. A análise de cluster requer a visualização da distribuição dos pontos de dados ou a aplicação de algoritmos específicos, o que não pode ser feito apenas com estatísticas descritivas e uma matriz de correlação.

No entanto, os dados apresentam características que sugerem que os agrupamentos podem existir e ser significativos:

1. **Transformação dos Dados (PCA):** As variáveis "V1" a "V28" exibem uma média muito próxima de zero e correlação praticamente nula entre si. Isso é uma forte indicação de que os dados originais foram transformados usando uma técnica como a Análise de Componentes Principais (PCA). O objetivo do PCA é, entre outros, maximizar a variância e criar eixos não correlacionados, o que pode ajudar a separar grupos distintos de dados.
2. **Variável "Class":** A coluna "Class" é tipicamente usada em datasets deste tipo para indicar uma categoria, como transações legítimas (0) e fraudulentas (1). Se houvesse dados para esta coluna (nas estatísticas fornecidas, a contagem é 0.0 e os valores são NaN), poderíamos inferir que o objetivo do dataset é justamente separar estas duas classes, que por definição formariam dois clusters distintos no espaço de características.
3. **Correlação com "Amount":** A variável "Amount" (valor da transação) possui correlações moderadas com algumas das variáveis "V". Por exemplo:
 - Correlação negativa com "V2" (-0.53).
 - Correlação positiva com "V7" (0.40).
 - Correlação positiva com "V20" (0.34).

Isso indica que o valor da transação está relacionado a alguns dos principais componentes, sugerindo que transações de valores diferentes podem se agrupar de maneiras distintas no espaço dimensional.

Em resumo, embora a estrutura dos dados sugira que eles foram preparados para uma tarefa de classificação (que implica a existência de grupos), as estatísticas fornecidas são insuficientes para provar a existência, a forma ou o número de clusters. A ausência de informações sobre a variável "Class" nos resumos estatísticos é a principal limitação para essa análise.

Pergunta 7: “Existem valores atípicos nos dados”?

8. Pergunta: Existem valores atípicos nos dados? (quero resposta objetiva)

Resposta: Sim, os dados apresentam valores atípicos (outliers) de forma clara, especialmente na coluna `Amount` (valor da transação).

A análise das estatísticas descritivas permite chegar a esta conclusão por meio de alguns indicadores principais:

1. **Diferença entre Média e Mediana:** A média da coluna é 88.35, enquanto a mediana (o valor do 50º percentil) é apenas 22.00. Uma média significativamente maior que a mediana indica que a distribuição dos dados é assimétrica à direita, ou seja, a média é "puxada" para cima por alguns valores muito altos, que são os outliers. `Amount`
2. **Discrepância entre o 3º Quartil e o Valor Máximo:** 75% de todas as transações têm um valor de até 77.16. No entanto, o valor máximo () registrado é 25.691,16. Essa diferença extrema confirma a existência de transações com valores muito superiores à grande maioria. `max`
3. **Análise nas Variáveis 'V':** O mesmo padrão pode ser observado, de forma menos intuitiva, nas variáveis (componentes principais). Por exemplo, na variável , o valor mínimo é -72.71, o que representa mais de 44 desvios-padrão (std=1.65) abaixo da média (próxima de zero), configurando um outlier extremo. `V "V2"`

Em resumo, a presença de valores máximos e mínimos que se distanciam drasticamente da maior parte dos dados, evidenciada pela diferença entre média/mediana e pelos extremos da distribuição, confirma a existência de outliers.

Pergunta 8: “Qual é a distribuição da variável V1”?



PROPOSTA DE PROJETO PARA O CURSO DE AGENTES AUTÔNOMOS COM REDES GENERATIVAS DO I2A2

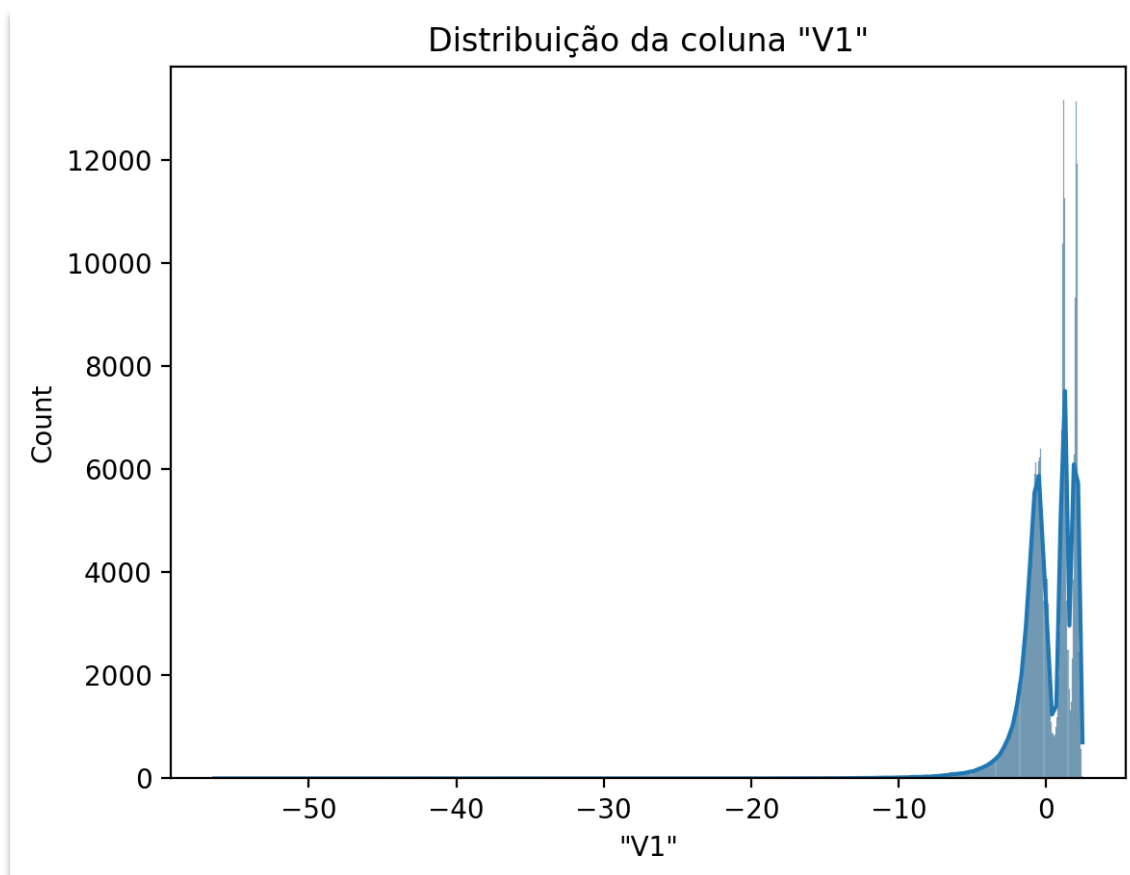


9. Pergunta: Qual é a distribuição da variável V1

Resposta: Com base nos dados fornecidos, a distribuição da variável V1 pode ser caracterizada da seguinte forma:

- **Centralidade:** A distribuição é centrada próximo de zero. A média é de (valor estatisticamente zero) e a mediana (50º percentil) é de . A proximidade entre a média e a mediana sugere que a parte central da distribuição é razoavelmente simétrica. `1.17e-15 0.018`
- **Dispersão:** O desvio padrão é de . O intervalo interquartil (a distância entre o 25º e o 75º percentil) vai de a , o que significa que 50% dos dados se concentram nesse intervalo. `1.96 -0.92 1.32`
- **Amplitude e Outliers:** A variável apresenta uma grande amplitude, com um valor mínimo de e um máximo de . O valor mínimo é muito mais distante da média do que o valor máximo, indicando a presença de outliers ou uma cauda longa no lado negativo da distribuição. `-56.41 2.45`

Em resumo, V1 é uma variável numérica com distribuição majoritariamente simétrica em torno de zero, mas com a presença de valores extremos (outliers), principalmente negativos.







10. CONCLUSÕES DO AGENTE


Após a realização das perguntas e análises interativas, o agente **EDAna** foi capaz de gerar **insights automáticos**, sintetizando os principais padrões observados nos dados. Essas conclusões oferecem uma **visão resumida e estratégica**, construída com base no histórico de interações e nas variáveis analisadas ao longo da sessão:

Pergunta: “Quais conclusões a EDAna obteve com base nas análises realizadas”?

Resposta:

 **Análise Interativa**

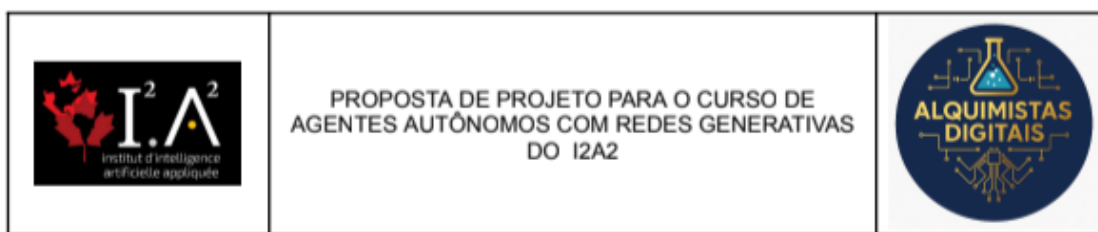
 **Conclusões da EDAna**



Conclusões automáticas

Aqui estão os insights que a EDAna identificou com base nas suas perguntas:

- Os valores variam dentro de um intervalo amplo.
- Foi observada uma variação temporal nas transações.
- Algumas variáveis apresentam valores recorrentes.
- A distribuição dos dados mostra concentração em faixas específicas.



11. CÓDIGOS FONTES UTILIZADOS

Os arquivos utilizados no desenvolvimento do agente **EDAna** foram:

- **app.py**: responsável pela interface Streamlit, controle de sessão, upload de arquivos, geração de gráficos e interação com o usuário;
- **llm_utils.py**: responsável pela comunicação com o modelo Gemini, geração de prompts e retorno das respostas;
- **agente.eda.py**: responsável por gerar conclusões automáticas com base no histórico de perguntas;
- **.env**: arquivo oculto que armazena a chave da API do Gemini, garantindo segurança e privacidade. Este arquivo não é incluído no relatório nem no repositório público.

Todos os arquivos do projeto foram devidamente organizados em uma estrutura única e compactados em formato .zip para fins de envio.

As chaves de acesso utilizadas, como a da API do Gemini, foram ocultadas conforme as diretrizes de segurança estabelecidas no desafio.

Além disso, o projeto encontra-se disponível em repositório público para consulta e validação:

🔗 *GitHub* – *EDAna*: <https://github.com/amandapraca/EDAna>

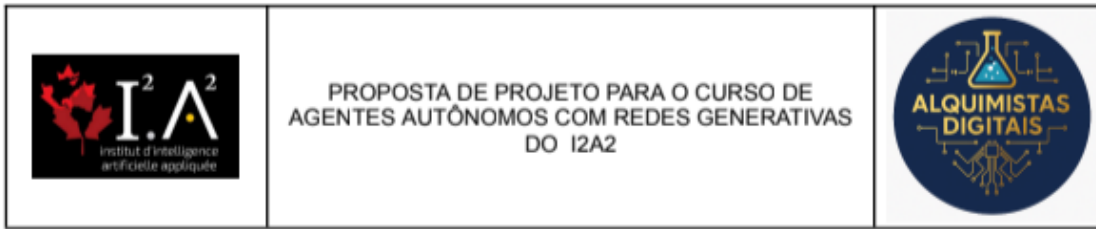
12. LINK DE ACESSO AO AGENTE

O agente **EDAna** foi desenvolvido para funcionar localmente, utilizando a plataforma **Streamlit** em ambiente Python. A aplicação é executada por meio do terminal, preferencialmente dentro do **Visual Studio Code**, onde o ambiente está corretamente configurado.

◆ Execução local

Para rodar o agente, é necessário:

- **Python 3.8+** instalado
- **Bibliotecas necessárias**: streamlit, pandas, matplotlib, seaborn, dotenv, google.generativeai, scikit-learn, numpy, python-dotenv.
- **Arquivo .env** contendo a chave da API Gemini



Comando para iniciar a aplicação principal:

```
streamlit run app.py
```

Comando para executar funções analíticas complementares:

```
python agente_eda.py
```

Observações:

- Recomenda-se executar os comandos dentro do terminal do **Visual Studio Code**, onde o ambiente está configurado corretamente.
- A interface será aberta automaticamente no navegador padrão após a execução do comando `streamlit run app.py`.
- O arquivo `agente_eda.py` pode ser utilizado para testes específicos ou análises complementares.

Link do agente no streamlit: <https://agente-eda-amanda.streamlit.app>

13. CONSIDERAÇÕES FINAIS

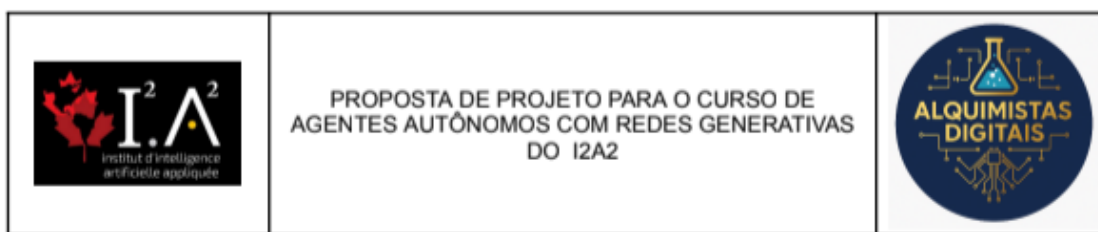
O projeto **EDAna** foi desenvolvido pela aluna Amanda Praça, como parte do Desafio 4 da disciplina *Agentes Autônomos com Redes Generativas*.

A aplicação foi testada com diferentes arquivos CSV, incluindo o conjunto de dados sobre **fraudes de cartão de crédito do Kaggle**, além de outros arquivos genéricos, demonstrando robustez na análise, geração de gráficos e respostas inteligentes com base nos dados fornecidos.

A interface foi cuidadosamente personalizada com **cores, ícones e mensagens de boas-vindas**, proporcionando uma experiência mais amigável, acessível e envolvente para o usuário.

A entrega é **individual e personalizada**, refletindo o estilo da desenvolvedora e incluindo uma assinatura no rodapé da aplicação:

Criado com carinho por Amanda Praça • Desafio I2A2 2025



Referências

CELSO. Professor. *Planilhas de apoio ao Desafio Relatório da Atividade Extra.* Instituto I2A2. Material fornecido diretamente pelo docente para fins de análise e desenvolvimento do projeto.

I2A2 ACADEMY. *Agentes Autônomos com Redes Generativas – Arquivos.* Disponível em: <https://sites.google.com/i2a2.academy/agentes-autonomos-com-ia-gen/arquivos>. Acesso em: 27 set. 2025.

MICROSOFT COPILOT. *Assistente IA – Dados Empresariais.* Plataforma de apoio ao desenvolvimento de agentes inteligentes. Disponível em: <http://localhost:5000/>. Acesso em: 27 set. 2025.

STREAMLIT. *Documentação oficial.* Disponível em: <https://docs.streamlit.io/>. Acesso em: 27 set. 2025.

GOOGLE AI. *Gemini API Documentation.* Disponível em: <https://ai.google.dev/>. Acesso em: 27 set. 2025.

PANDAS. *Pandas Documentation.* Disponível em: <https://pandas.pydata.org/docs/>. Acesso em: 27 set. 2025.

SEABORN. *Seaborn Documentation.* Disponível em: <https://seaborn.pydata.org/>. Acesso em: 27 set. 2025.