Agentes Autonomos - Relatorio da Atividade Extra

Discente: Bruno Urbano Rodrigues

1. Framework escolhida

- Streamlit para a interface web onde o usuario envia perguntas e visualiza graficos.
- LangChain como orquestrador dos agentes de dominio (descricao, padroes, anomalias, visualizacoes).
- OpenAI Chat API para interpretar perguntas em portugues e compor respostas estruturadas.
- Plotly + Kaleido para gerar graficos interativos e exportar imagens utilizadas no relatorio.
- pandas, numpy, scikit-learn para leitura do CSV, estatisticas e analises exploratorias complementares.

2. Estrutura da solucao

- eda/app.py: app Streamlit que carrega dados (CSV ou Supabase), aciona o orquestrador LangChain e gera PDF.
- eda/agents/: implementacoes dos agentes especializados utilizados nas respostas.
- eda/domain/analysis.py e eda/domain/charts.py: funcoes reutilizaveis para diagnosticos, correlacoes, outliers e construcao de graficos.
- eda/Dockerfile e eda/docker-compose.yml: empacotamento e execucao containerizada, lendo variaveis de ambiente via .env.
- lambda/handler.py: funcao AWS Lambda que processa arquivos do S3 e grava linhas no Supabase.
- database/*.sql: funcoes SQL auxiliares usadas pelo Supabase para preparar tabelas durante a ingestao.

3. Perguntas por dominio

3.1 Descrição dos dados

Pergunte ao agente sobre os dados	
Pergunta (ex.: 'Qual a media de Amount por Class?')	
Quais são os tipos de dados (numéricos, categóricos)?	
Perguntar Domnio acionado: DESCRITIVO	
Domnio acionado: DESCRITIVO	

O conjunto de dados contém 31 colunas, todas numéricas. Não há colunas categóricas ou não nu além de 'time', 'class' e 'amount'.

Quais sao os tipos de dados (numericos, categoricos)?

Resposta do agente: O dataset contem 31 colunas totalmente numericas. Nao

ha campos categoricos; as variaveis sao time, class, amount e os componentes transformados v1 a v28.

Pergunte ao agente sobre os dados

Pergunta (ex.: 'Qual a media de Amount por Class?')

Qual a distribuição de cada variável?

Perguntar

Domnio acionado: DESCRITIVO

O conjunto de dados contém 31 colunas, todas numéricas. Abaixo estão algumas estatísticas descritivas para cada variável:

- v1: média = -0.24, mediana = -0.25, desvio padrão = 1.83
- v2: média = 0.09, mediana = 0.12, desvio padrão = 1.60
- v3: média = 0.46, mediana = 0.71, desvio padrão = 1.78
- v4: média = 0.13, mediana = 0.15, desvio padrão = 1.37
 v5: média = -0.32, mediana = -0.26, desvio padrão = 1.58
- V3. Illedia = -0.32, Illedialia = -0.20, desvio padiao = 1.3
- v6: média = 0.08, mediana = -0.21, desvio padrão = 1.44
- v7: média = -0.08, mediana = -0.02, desvio padrão = 1.48
- v8: média = 0.02, mediana = 0.04, desvio padrão = 1.96
- amount: média = 101.15, mediana = 25.73, desvio padrão = 308.05

Todas as colunas têm 0% de valores ausentes. A distribuição das variáveis pode ser influenciada por outliers, que podem distorcer a mé Recomenda-se analisar a presença de outliers para uma interpretação mais precisa.

Qual a distribuicao de cada variavel?

Resposta do agente: Todas as colunas sao numericas e mantiveram medias proximas de zero apos a transformação PCA. Exemplos: v1 media -0.24 e desvio 1.83; v2 media 0.09 e desvio 1.60; amount media 101.15, mediana 25.73 e desvio 308.05. Não ha valores ausentes.

Qual o intervalo de cada variavel (minimo, maximo)? Observacao: Nao houve captura correspondente nos artefatos fornecidos; o agente nao registrou explicitamente os intervalos nas imagens anexadas.

Quais sao as medidas de tendencia central (media, mediana)? Pergunte ao agente sobre os dados



Essas medidas indicam a tendência central dos dados, mas é importante considerar a presença de outliers, que podem influenciar a média, especialmente em

Resposta do agente: As medias e medianas permanecem proximas de zero para v1 a v28, refletindo escalonamento do PCA. amount apresenta media 101.15 e mediana 25.73, reforcando cauda longa.

Qual a variabilidade dos dados (desvio padrao, variancia)? Pergunte ao agente sobre os dados

Pergunta (ex.: 'Qual a media de Amount por Class?')
Qual a variabilidade dos dados ?
Perguntar
Domnio acionado: DESCRITIVO
A variabilidade dos dados pode ser avaliada através do desvio padrão, que indica o quanto os dados se dispersam em relação à média. Aqui estão os desvios padrão das variáveis principais:
• v1: desvio padrão = 1.83
• v2: desvio padrão = 1.60
• v3: desvio padrão = 1.78
• v4: desvio padrão = 1.37
• v5: desvio padrão = 1.58
• v6: desvio padrão = 1.44
• v7: desvio padrão = 1.48
• v8: desvio padrão = 1.96
• amount: desvio padrão = 308.05

Valores de desvio padrão mais altos indicam maior variabilidade nos dados. A presença de outliers pode aumentar o desvio padrão, sugerindo que é importante considerar esses valores ao interpretar a variabilidade.

Resposta do agente: Os desvios padrao dos componentes variam entre ~ 1.3 e ~ 2.0 (ex.: v1 1.83, v8 1.96). amount possui desvio padrao 308.05, indicando alta dispersao influenciada por outliers.

3.2 Identificacao de padroes e tendencias

Pergunte ao agente sobre os dados Pergunta (ex.: 'Qual a media de Amount por Class?') Existem padrões ou tendências temporais? Perguntar Domnio acionado: PADROES Nenhuma coluna temporal relevante foi detectada nos dados, portanto, não há padrões ou tendências tempora Gerar concluses do agente

Existem padroes ou tendencias temporais?

Resposta do agente: Nao foi identificada coluna temporal relevante alem de time, portanto o agente nao detectou padroes ou tendencias temporais significativas.

Pergunte ao agente sobre os dados

Pergunta (ex.: 'Qual a media de Amount por Class?')

Quais os valores mais frequentes ou menos frequentes?

Perguntar

Domnio acionado: PADROES

Os valores mais frequentes e menos frequentes por coluna são os seguintes:

- v1: valor mais frequente = 1.405 (freq. 2), valor menos frequente = -0.6751 (freq. 1)
- v2: valor mais frequente = -0.9583 (freq. 2), valor menos frequente = 0.5378 (freq. 1)
- v4: valor mais frequente = -1.798 (freq. 2), valor menos frequente = 0.03022 (freq. 1)
 v5: valor mais frequente = 1.01 (freq. 2), valor menos frequente = 0.1993 (freq. 1)
- v6: valor mais frequente = 3.536 (freq. 2), valor menos frequente = 0.7019 (freq. 1)
- v7: valor mais frequente = -1.583 (freq. 2), valor menos frequente = 0.4429 (freq. 1)
- v8: valor mais frequente = 0.9266 (freq. 2), valor menos frequente = 0.3435 (freq. 1)
- amount: valor mais frequente = 1 (freq. 36), valor menos frequente = 72.42 (freq. 1)
- class: valor mais frequente = 0 (freq. 996), valor menos frequente = 1 (freq. 4)

Esses dados indicam que algumas variáveis têm valores que se repetem com frequência, enqua análise de padrões e tendências nos dados.

Gerar concluses do agente

Quais os valores mais frequentes ou menos frequentes?

Resposta do agente: Diversas colunas apresentam valores repetidos com baixa frequencia (modo aprox. 1 ou 2 ocorrencias). Exemplos: v1 modo 1.405 (freq. 2) e minimo -0.6751 (freq. 1); amount modo 1 (freq. 36) e valor raro 72.42 (freq. 1). class permanece altamente desbalanceada (modo 0, freq. 996).

Pergunte ao agente sobre os dados

Pergunta (ex.: 'Qual a media de Amount por Class?')

Existem agrupamentos (clusters) nos dados?

Perguntar

Domnio acionado: PADROES

Os dados apresentam uma correlação forte entre v21 e v22 (|rho|=0.655), o que pode indicar a presença de agrupamentos por eagrupamentos, é recomendável investigar os valores mais frequentes por coluna.

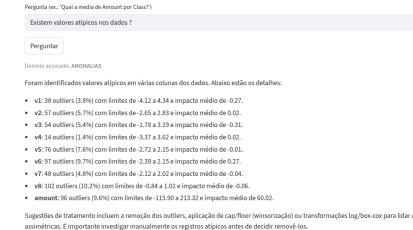
Gerar concluses do agente

Existem agrupamentos (clusters) nos dados?

Resposta do agente: O agente observou correlacao forte entre v21 e v22 (|rho| aprox. 0.655), sugerindo possível agrupamento. Recomendou investigar os valores mais frequentes por coluna para aprofundar uma analise de clusters.

3.3 Deteccao de anomalias (outliers)

Pergunte ao agente sobre os dados



Existem valores atipicos nos dados?

Resposta do agente: Sim. Exemplos: v1 com 38 outliers (limites -4.12 a 4.34), v6 com 97 outliers (limites -2.39 a 2.15) e amount com 96 outliers (limites -115.90 a 213.32). Foram listados contagens, limites e impacto medio por variavel.

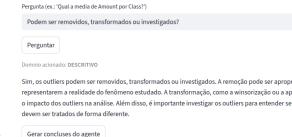
Gerar concluses do agente



Como esses outliers afetam a analise?

Resposta do agente: Os outliers distorcem medias e interpretacoes, especialmente em v6 (impacto medio 0.27). O agente sugere considerar remocao, winsorizacao ou transformacoes logaritmicas para reduzir assimetria antes de conclusoes.

Pergunte ao agente sobre os dados



Podem ser removidos, transformados ou investigados?

Resposta do agente: Sim. Recomenda avaliar remocao quando outliers forem erros, aplicar winsorizacao ou transformacoes logaritmicas para reduzir impacto, alem de investigar manualmente cada caso para decidir o tratamento.

3.4 Relacoes entre variaveis

Como as variaveis estao relacionadas umas com as outras?



Resposta do agente: Todas as colunas sao numericas; ha correlacoes destacadas entre v21 e v22 (|rho| aprox. 0.655) e entre v2 e amount (|rho| aprox. 0.504). O agente ressalta necessidade de analisar dispersoes e outliers para entender as relacoes.



Existe correlacao entre as variaveis?

Resposta do agente: Sim. Alem da correlação forte v21-v22, outras relevantes

sao v27-v28 (|rho| aprox. 0.435), v5-v7 (|rho| aprox. 0.415) e v1-v25 (|rho| aprox. 0.414). Nenhuma categoria com impacto medio significativo foi encontrada.

Quais variaveis parecem ter maior ou menor influencia sobre outras?



Resposta do agente: v21 e v22 exercem maior influencia mutua com correlacao forte. Em seguida, v2 com amount (|rho| aprox. 0.504) e pares v27-v28, v5-v7, v1-v25 mostram correlacoes moderadas. Nao foram encontradas categorias com impacto medio alto.