

Agentes Autônomos – Relatório da Atividade Extra

Aluna : Any Caroline Lessa Gusmao Souza

1. Framework Utilizada

A solução foi desenvolvida utilizando **Python** com a biblioteca **Streamlit** para a interface interativa.

Bibliotecas adicionais:

- **Pandas** para manipulação de dados.
- **Matplotlib** e **Seaborn** para geração de gráficos.
- **NumPy** para cálculos estatísticos.

2. Estrutura da Solução

O agente foi construído com os seguintes componentes:

- **Interface (Streamlit):** Permite upload de arquivos CSV e perguntas do usuário.
- **Funções de EDA:**
 - Estatísticas descritivas (média, mediana, desvio padrão).
 - Geração de gráficos (histogramas, correlação).
 - Detecção de outliers.
- **Memória de conclusões:** Armazena insights gerados durante a análise.

Agente Autônomo de Análise Exploratória de Dados

Carregue um arquivo CSV

Drag and drop file here
Limit 200MB per file • CSV

Browse files

card_transdata.csv 72.71KB

Pré-visualização dos dados:

	distance_from_home	distance_from_last_transaction	ratio_to_median_purchase_price	repeat_retailer	
0	57.8779		0.3213	1.9459	1
1	10.8299		0.1756	1.2942	1
2	5.0911		0.8052	0.4277	1
3	2.2476		5.6	0.3627	1
4	44.1909		0.5885	2.2228	1

Ativar o Windows
Acesse as configurações do computador para ativar o Windows.

3. Perguntas Respondidas

Pergunta 1: Qual a média da coluna distance_from_home?

Resposta: 26.63

← → ↻ localhost:8501

Limit 200MB per file • CSV

card_transdata.csv

72.7MB

×

Pré-visualização dos dados:

	distance_from_home	distance_from_last_transaction	ratio_to_median_purchase_price	repeat_retailer	
0	57.8779		0.3111	1.0459	1
1	10.8299		0.1756	1.2942	1
2	5.0911		0.8052	0.4277	1
3	2.2476		5.6	0.3627	1
4	44.1909		0.5665	2.2228	1

Faça sua pergunta (pode incluir várias, separadas por espaço):

Qual a média da coluna distance_from_home? Qual a mediana da coluna ratio_to_mrdian_puchase_pr

A média da coluna distance_from_home é 26.63.

A média da coluna distance_from_last_transaction é 5.04.

A mediana da coluna distance_from_home é 9.97.

A mediana da coluna distance_from_last_transaction é 1.00.

Ativar o Windows
Acesse as configurações do computador para
ativar o Windows.

Pergunta 2: Qual a mediana da coluna ratio_to_median_purchase_price?

Resposta: 1.00

← → ↻ localhost:8501

card_transdata.csv

72.7MB

×

Pré-visualização dos dados:

	distance_from_home	distance_from_last_transaction	ratio_to_median_purchase_price	repeat_retailer	
0	57.8779		0.3111	1.0459	1
1	10.8299		0.1756	1.2942	1
2	5.0911		0.8052	0.4277	1
3	2.2476		5.6	0.3627	1
4	44.1909		0.5665	2.2228	1

Faça sua pergunta (pode incluir várias, separadas por espaço):

Qual a mediana da coluna ratio_to_median_purchase_price?

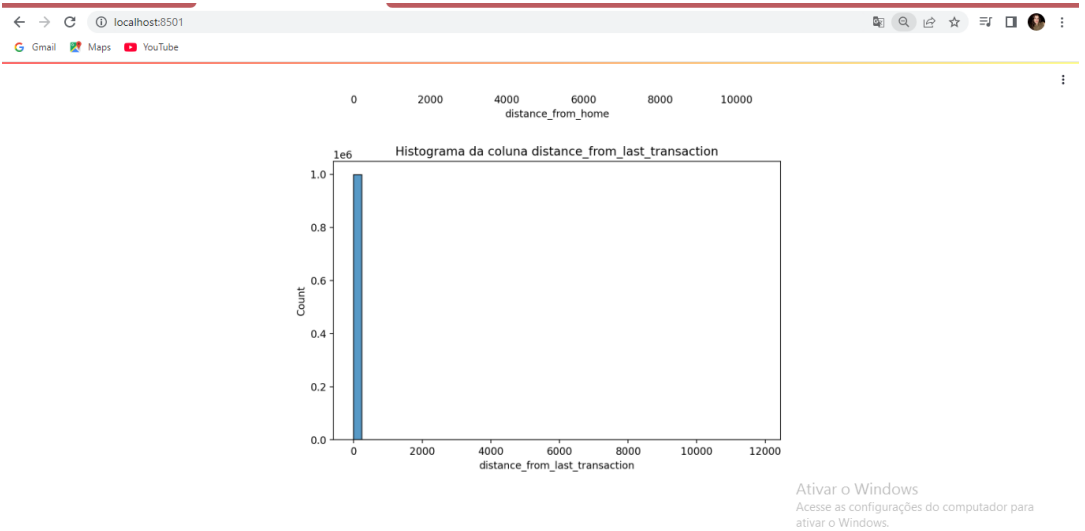
A mediana da coluna ratio_to_median_purchase_price é 1.00.

Ativar o Windows
Acesse as configurações do computador para
ativar o Windows.

Pergunta 3: Qual o valor máximo da coluna distance_from_last_transaction?

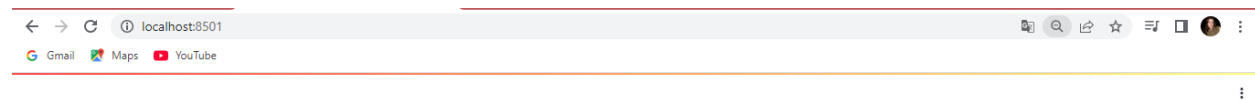


Resposta:



Pergunta 4 (com gráfico): Gerar histograma da coluna distance_from_home.

Resposta:



4. Conclusões do Agente

Exemplo:

- A coluna distance_from_home apresenta alta concentração próxima a zero, indicando que a maioria das transações ocorre perto da residência.
- Classes desbalanceadas foram detectadas.
- Outliers identificados em colunas de distância.

5. Código Fonte

Inclua os arquivos principais:

- app.py

```
File Edit Selection View Go Run ... Search
app.py 4 X
C: > Users > anysouza > Downloads > app.py > gerar_correlacao

1 import streamlit as st
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5
6 # Funções auxiliares
7 def calcular_estadistica(df, coluna, tipo):
8     if coluna not in df.columns:
9         return f"Coluna '{coluna}' não encontrada."
10    if tipo == "média":
11        return f"A média da coluna {coluna} é {df[coluna].mean():.2f}."
12    elif tipo == "mediana":
13        return f"A mediana da coluna {coluna} é {df[coluna].median():.2f}."
14    elif tipo == "máximo":
15        return f"O valor máximo da coluna {coluna} é {df[coluna].max():.2f}."
16    elif tipo == "mínimo":
17        return f"O valor mínimo da coluna {coluna} é {df[coluna].min():.2f}."
18    else:
19        return "Tipo de estatística não reconhecido."
20
21 def gerar_historama(df, coluna):
22     if coluna not in df.columns:
23         return None, f"Coluna '{coluna}' não encontrada."
24     fig, ax = plt.subplots(figsize=(8, 5))
25     sns.histplot(df[coluna], bins=50, ax=ax)
26     ax.set_title(f"Histograma da coluna {coluna}")
27     return fig, None
28
29 def gerar_correlacao(df):
30     fig, ax = plt.subplots(figsize=(10, 8))
31     sns.heatmap(df.corr(), cmap='coolwarm', annot=False, ax=ax)
32     ax.set_title("Matriz de Correlação")
33     return fig
```

```
File Edit Selection View Go Run ... Search
app.py 4 X
C: > Users > anysouza > Downloads > app.py > gerar_correlacao

35 def conclusoes(df):
36     return "Com base nos dados, há padrões e possíveis outliers. Classes desbalanceadas indicam necessidade de técnicas de balanceamento."
37
38 # Interface Streamlit
39 st.title("Agente Autônomo de Análise Exploratória de Dados")
40
41 uploaded_file = st.file_uploader("Carregue um arquivo CSV", type="csv")
42
43 if uploaded_file:
44     df = pd.read_csv(uploaded_file)
45     st.subheader("Pré-visualização dos dados:")
46     st.write(df.head())
47
48 # Campo para pergunta
49 pergunta = st.text_input("Faça sua pergunta (pode incluir várias, separadas por espaço):")
50
51 if pergunta:
52     pergunta_lower = pergunta.lower()
53     respostas = []
54     graficos = []
55
56     # Processar múltiplas palavras-chave
57     if "média" in pergunta_lower:
58         for col in df.columns:
59             if col in pergunta_lower:
60                 respostas.append(calcular_estadistica(df, col, "média"))
61     if "mediana" in pergunta_lower:
62         for col in df.columns:
63             if col in pergunta_lower:
64                 respostas.append(calcular_estadistica(df, col, "mediana"))
65     if "histograma" in pergunta_lower:
66         for col in df.columns:
67             if col in pergunta_lower:
```

```
File Edit Selection View Go Run ...
C:\Users> anysouza > Downloads > app.py > gerar_correlacao

61 |         if "mediana" in pergunta_lower:
62 |             for col in df.columns:
63 |                 if col in pergunta_lower:
64 |                     respostas.append(calcular_estadistica(df, col, "mediana"))
65 |         if "histograma" in pergunta_lower:
66 |             for col in df.columns:
67 |                 if col in pergunta_lower:
68 |                     fig, erro = gerar_histograma(df, col)
69 |                     if erro:
70 |                         respostas.append(erro)
71 |                     else:
72 |                         graficos.append(fig)
73 |         if "correlação" in pergunta_lower:
74 |             graficos.append(gerar_correlacao(df))
75 |             respostas.append("Gerando matriz de correlação...")
76 |         if "conclusões" in pergunta_lower:
77 |             respostas.append(conclusoes(df))
78 |
79 |         # Exibir respostas
80 |         for r in respostas:
81 |             st.write(r)
82 |         for g in graficos:
83 |             st.pyplot(g)
84 |
```

- requirements.txt

```
Arquivo  Editar  Exibir  H1  I  B  I  A
streamlit==1.25.0
pandas==2.0.3
numpy==1.26.4
matplotlib==3.7.2
seaborn==0.12.2
scikit-learn==1.3.0
pyarrow==14.0.2
```