

Simplifica Python



APOSTILA

ANÁLISE DE DADOS COM PYTHON



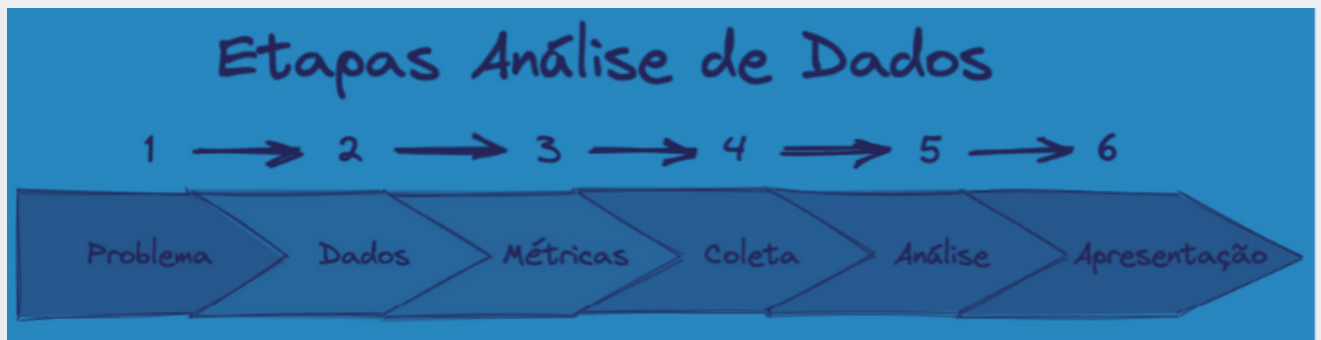
Alexandre Sauer

@prof.alesauer

• Especialista em Python
• Mestre em Sistemas da Informação

ANÁLISE DE DADOS

A IMPORTÂNCIA DA ANÁLISE DE DADOS E A UTILIZAÇÃO DO PANDAS NA GESTÃO FAMILIAR



Em um mercado de trabalho competitivo, diferenciar-se é essencial. A análise de dados com Python é uma habilidade poderosa que pode elevar sua carreira e destacar você da concorrência. Vamos entender as etapas essenciais da análise de dados e como cada uma delas contribui para sua diferenciação.

Etapas da Análise de Dados

- **Problema:**

- Descrição: Identificar e definir claramente o problema que precisa ser resolvido é a primeira e mais crucial etapa. Um entendimento preciso do problema direciona todo o processo de análise de dados.
- Exemplo: "Como podemos aumentar a retenção de clientes em nossa plataforma?"

ANÁLISE DE DADOS

- **Dados:**

- Descrição: Coletar os dados relevantes para o problema. Isso envolve identificar quais dados são necessários e onde eles podem ser obtidos.
- Exemplo: Dados de comportamento do usuário, histórico de compras, feedbacks de clientes.

- **Métricas:**

- Descrição: Definir as métricas que serão usadas para medir o sucesso da solução. Métricas claras e relevantes ajudam a avaliar o impacto das ações tomadas.
- Exemplo: Taxa de retenção de clientes, taxa de churn, valor de vida do cliente (LTV).

- **Coleta:**

- Descrição: Reunir os dados necessários de diferentes fontes, garantindo que sejam precisos e relevantes. A coleta pode envolver extração de bases de dados, APIs, surveys, entre outros.
- Exemplo: Extração de dados do sistema CRM, coleta de dados via Google Analytics.



ANÁLISE DE DADOS

- **Análise:**

- Descrição: Processar e analisar os dados coletados para descobrir padrões, insights e tendências. Essa etapa envolve técnicas estatísticas e de machine learning.
- Exemplo: Análise de segmentação de clientes, identificação de padrões de compra.

- **Apresentação:**

- Descrição: Apresentar os resultados da análise de maneira clara e compreensível. Visualizações e relatórios são fundamentais para comunicar os insights e ajudar na tomada de decisões.
- Exemplo: Relatórios de dashboards interativos, apresentações visuais com gráficos e tabelas.

NOSSO PROJETO

GESTÃO FINANCEIRA DE UM CASAL

Cenário:

Ana e João são um casal que compartilha uma conta conjunta de cartão de crédito. Eles têm percebido que, no final do mês, nunca sabem exatamente onde o dinheiro foi gasto. As faturas chegam e, com tantas transações, é difícil ter uma visão clara de suas despesas. Eles querem entender melhor seus hábitos de consumo e identificar onde podem economizar.

NOSSO PROJETO

Objetivo do Projeto:

Este projeto realiza uma análise detalhada de transações de cartões de crédito, utilizando uma abordagem baseada em dados. Inicialmente, ele carrega os dados de um arquivo CSV contendo informações sobre os gastos dos titulares, como o nome do comprador, categoria da compra, data, valor e estabelecimento.



DADOS DO CARTÃO DE CRÉDITO

CONTEXTO: GESTÃO FINANCEIRA DE UM CASAL

ARQUIVO: fatura.csv

Este arquivo contém informações detalhadas sobre as transações de cartão de crédito de um casal, Ana e João. Eles compartilham uma conta conjunta e querem entender melhor onde estão gastando seu dinheiro. A análise desses dados pode fornecer insights valiosos sobre seus hábitos de consumo e ajudar a identificar áreas onde podem economizar.

DADOS DO CARTÃO DE CRÉDITO

CONTEXTO: GESTÃO FINANCEIRA DE UM CASAL

Estrutura do Arquivo

O arquivo possui as seguintes colunas:

- NumeroCartao: Número do cartão de crédito utilizado na transação.
- data_compra: Data em que a compra foi realizada.
- hora_compra: Hora em que a compra foi realizada.
- estabelecimento: Tipo de estabelecimento onde a compra foi feita.
- categoria: Categoria da despesa (e.g., Alimentação, Transporte, Lazer).
- ValorCompra: Valor da compra.

```

1 NumeroCartao,data_compra,hora_compra,estabelecimento,categoria,ValorCompra
2 7840517590708225,2024-05-16,08:15:06,Lazer,Parque de Diversões,898.93
3 7142474765870246,2024-05-02,01:31:42,Alimentação,Restaurante,795.62
4 7142474765870246,2024-05-07,16:49:44,Viagem,Hotel,718.54
5 7840517590708225,2024-05-10,23:37:25,Eletrônicos,Loja de Eletrônicos,344.81
6 7142474765870246,2024-05-15,08:49:21,Mercado,Mercado,815.07
7 7142474765870246,2024-05-06,14:11:14,Farmácia,Farmácia,822.05
8 7840517590708225,2024-05-01,13:05:24,Lazer,Cinema,35.26
9 7142474765870246,2024-05-07,11:34:44,Mercado,Supermercado,861.8
10 7840517590708225,2024-04-27,04:35:33,Alimentação,Padaria,761.31
11 7840517590708225,2024-05-05,05:28:58,Transporte,Posto de Gasolina,195.44
12 7142474765870246,2024-05-06,11:08:42,Vestuário,Boutique,753.23
13 7142474765870246,2024-05-03,19:44:09,Lazer,Cinema,766.82
14 7142474765870246,2024-05-14,02:28:25,Farmácia,Drogaria,98.37
15 7840517590708225,2024-05-14,18:09:10,Viagem,Agência de Viagens,330.48
16 7142474765870246,2024-04-30,15:39:32,Eletrônicos,Tecnologia,141.91
17 7840517590708225,2024-04-19,13:50:07,Viagem,Agência de Viagens,630.4
18 7840517590708225,2024-05-12,02:58:08,Lazer,Cinema,260.62
19 7142474765870246,2024-05-08,19:07:08,Alimentação,Restaurante,406.89
20 7142474765870246,2024-05-10,14:34:24,Transporte,Estacionamento,143.86
21 7840517590708225,2024-04-25,00:50:18,Alimentação,Cafeteria,532.68
22 7840517590708225,2024-04-27,18:13:51,Eletrônicos,Tecnologia,655.4
23 7142474765870246,2024-04-21,15:27:46,Transporte,Estacionamento,819.97
24 7142474765870246,2024-05-06,13:07:00,Mercado,Hipermercado,334.72
25 7142474765870246,2024-04-17,14:41:15,Transporte,Estacionamento,221.72
26 7840517590708225,2024-04-28,12:12:02,Mercado,Mercado,289.62
27 7840517590708225,2024-05-16,03:59:57,Alimentação,Restaurante,20.13
28 7142474765870246,2024-05-08,08:30:37,Vestuário,Loja de Roupas,961.54
29 7840517590708225,2024-05-03,14:32:43,Transporte,Posto de Gasolina,899.05
30 7142474765870246,2024-05-01,19:34:36,Farmácia,Farmácia,12.53
  
```


SOBRE O CÓDIGO

Este código utiliza bibliotecas populares do **Python**, como **pandas** para manipulação de dados e **matplotlib** para geração de gráficos. Ele tem como objetivo analisar os gastos de titulares de cartões de crédito e apresentar os resultados de forma visual.

Após carregar os dados, são realizadas diversas operações de agregação. A primeira delas calcula o valor total gasto por todos os titulares. Em seguida, o código organiza os gastos de forma a somar o valor total de compras por titular, apresentando uma visão clara de quem gastou mais.

Além disso, o código também agrupa os dados por categoria de compra, permitindo entender em quais tipos de serviços ou produtos os titulares gastaram mais dinheiro, como alimentação, lazer, vestuário, entre outros. Outro ponto de análise é a distribuição dos gastos ao longo do tempo, mostrando os dias em que os gastos foram mais altos.



SOBRE O CÓDIGO

Para facilitar a interpretação dos dados, o código gera vários tipos de gráficos: um gráfico de barras que ilustra o valor gasto por cada titular, um gráfico de barras horizontal para comparar as categorias de gastos, um gráfico de pizza que demonstra a proporção dos gastos entre os titulares, e um gráfico de linha que exhibe como os gastos variaram ao longo do tempo.

Essas visualizações são cruciais para entender os padrões de consumo, identificar os maiores gastadores e reconhecer as categorias com maior impacto no orçamento.

Para ver o código completo e interagir com ele, **[acesse o material da aula.](#)**

