

Projeto III

Análise Exploratória de Dados Utilizando o Python ou RStudio

FATEC Rubens Lara | Professor Jobel Corrêa. MBA, MsC. | 23/03/2023

I.1 Base de dados

Utilizaremos a base de dados “Varejo.csv” disponibilizada na pasta do curso no endereço <https://abre.ai/fWDK>. Daqui em diante essa base será chamada de Varejo para facilitar a compreensão.

I.1.1 Descrição da base de dados

A base de dados Varejo é um recorte de 50.000 compras de 1.000 clientes realizadas em uma rede de supermercados no período de 2010 à 2022.

I.1.2 Lista de campos existentes na base de dados

A Figura I.1 mostra um resumo da base de dados Varejo, com a descrição de cada campo a seguir.

DATA	CO_ID	CL_ID	CO_ID	CL_GENERO	CL_EC	CL_FHL	CL_SEG	PR_ID	PR_CAT	PR_NOME
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800210	ALIMENTOS	BANANA
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800216	ALIMENTOS	CAFE
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800247	ALIMENTOS	PAO DE FORMA
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800225	ALIMENTOS	DOCE
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800248	ALIMENTOS	PAPINHA INFANTIL
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	600402	HIGIENE	CREME
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	400602	PET	RACAO SECA PARA GATOS
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	600417	HIGIENE	REPELENTE
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800245	ALIMENTOS	OVOS
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800222	ALIMENTOS	CORACAO DE FRANGO
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800252	ALIMENTOS	QUEIJO DE CABRA
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800243	ALIMENTOS	MOLHO BARBECUE
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	500521	LIMPEZA	VASSOURA
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800233	ALIMENTOS	LATA DE MILHO
02/01/2010	10872	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	600408	HIGIENE	GEL
03/01/2010	11059	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800248	ALIMENTOS	PAPINHA INFANTIL
03/01/2010	11059	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	600415	HIGIENE	PRESERVATIVO
03/01/2010	11059	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	600419	HIGIENE	SHAMPOO
03/01/2010	11059	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	600402	HIGIENE	CREME
03/01/2010	11059	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	800259	ALIMENTOS	SNACKS
03/01/2010	11059	1	VERDADEIRO	M	CASADO OU UNIAO ESTAVEL	0	C	500518	LIMPEZA	SABAO EM PO

Figura I.1: Visão compacta da base de dados Varejo.csv | Fonte: Autor, adaptado da base de dados

1. DATA: Data da compra;
2. CO_ID: Identificação do número de compra (número da nota fiscal);
3. CL_ID: Identificação do cliente (número do cliente);
4. CL_GENERO: Sexo biológico informado pelo cliente;

5. CL_EC: Estado civil do cliente;
6. CL_FHL: Número de filhos do cliente;
7. CL_SEG: Segmentação econômica do cliente (classe A, B ou C);
8. PR_ID: Código do produto (SKU) adquirido;
9. PR_CAT: Categoria do produto adquirido;
10. PR_NOME: Nome do produto adquirido.

I.1.3 Análise Exploratória de Dados (AED)

Utilizando o ambiente Google Colab com a linguagem Python ou R, com auxílio dos e-books disponibilizados na pasta do curso e da ferramenta ChatGPT você deve realizar:

- a) Análise Exploratória de Dados (AED) dos dados da base Varejo. Faça um sumário dos dados, realize correlações, verifique a necessidade de mudança de variáveis, construa histogramas e bloxplots, monte gráficos de dispersão de dados.
- b) Apresentação em PowerPoint ou Google Presentation com o storytelling adequado.