

Apresentação da Base de Dados

Problema: Detecção de anomalias de preços (Americanas)



Tabela ofertas_caneta.csv

- **Tamanho tabela:** 5 milhões e 500 mil entradas, 7 colunas

- **Janela Temporal:**

Formato: milissegundos

Máximo: 2022-01-16 04:56:39.084000

Mínimo: 2021-06-01 00:06:22.654000

1	id_seller	id_oferta	id_item_pai	id_item_filho	preco_da_oferta	timestamp_evento	id_evento
2	Cf1BN5s/z	ZazDvXLFp	TYlo9zZLIghY	Czfb5udg99xoF	27	2021-06-01 00:06:22.654 UTC	4fb5c9959d6
3	Cf1BN5s/z	ZazDvXLFp	TYlo9zZLIghY	Czfb5udg99xoF	27	2021-06-01 00:06:22.773 UTC	04ea0bd78a
4	Cf1BN5s/z	JKxXAxQN	3wH7C6aytlC	3V/J0RQjokDdc	77	2021-06-01 00:07:25.367 UTC	f0552e03d40
5	Cf1BN5s/z	JKxXAxQN	3wH7C6aytlC	3V/J0RQjokDdc	77	2021-06-01 00:07:25.549 UTC	e03e846edc
6	Cf1BN5s/z	qwUA82s+	zhidga+XmPc	1pokPB4i8fEMi	27	2021-06-01 00:07:56.516 UTC	4d054b905c
7	Cf1BN5s/z	qwUA82s+	zhidga+XmPc	1pokPB4i8fEMi	27	2021-06-01 00:07:56.637 UTC	43a547829d
8	c7P8w2ZP	s71aiy61S	Y4woXDinZ6	0QDW3S357yKl	367.5	2021-06-01 00:08:02.661 UTC	28ba969f20c
9	c7P8w2ZP	3iIT2AeU2	gD5ga85Eaiq	83QoAxkGZr09	262.5	2021-06-01 00:08:48.62 UTC	ec4079e81a

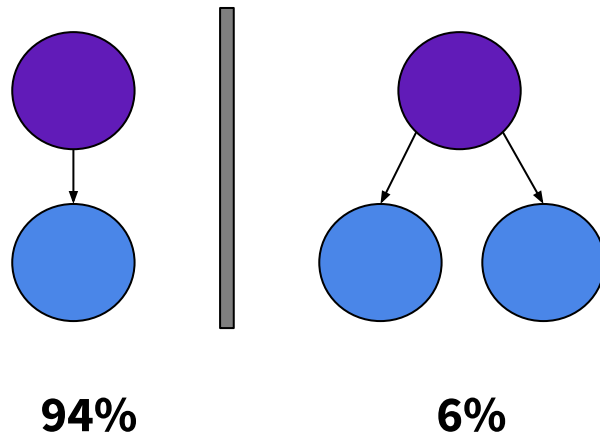
Tabela ofertas_caneta.csv: Qtde de produtos

- **Produtos_pai na base:**

123490

- **Produtos_filho na base:**

132991



- Ou seja, somente $132490 - 123490 = 9000$ **produtos_filho** pertencem a **mais de um produto_pai**. Isto é, **6,77%** de todos os produtos_filhos.

Tabela ofertas_caneta.csv: Qtde de Ofertas

- Quantidade de ofertas distintas realizadas:

144266

- Quantidade de Produtos_filho na base:

132991

Logo, **92%** das ofertas se referem ao **diferentes produto_filho!**

- Quantidade de Produtos_pai na base:

123490

Logo, **86%** das ofertas se referem ao **diferentes produto_pai!**

