O [crosstab](https://pandas.pydata.org/docs/reference/api/pandas.crosstab.html), ou tabela cruzada, é uma função da biblioteca pandas utilizada para analisar a relação entre duas variáveis categóricas. Ela cria uma tabela com as variáveis categóricas como índices e colunas onde são exibidos os valores únicos de cada variável nesse formato. Os valores da tabela são a combinação entre as duas variáveis.

Ela é útil para fins de comparação entre as frequências de diferentes grupos, identificação de padrões ou associações entre variáveis.

A sintaxe básica do crosstab é a seguinte:

pd.crosstab(index, columns, values, aggfunc, normalize)

Onde:

* **index:** valores para agrupar nas linhas da tabela
* **columns:** valores para agrupar nas colunas da tabela
* **values:** valores da tabela a partir da função de agregação (aggfunc)
* **aggfunc:** função de agregação dos dados, por exemplo ”sum”
* **normalize:** normalização dos dados. Pode ser por colunas, linhas ou ambos.

Aqui abaixo temos um exemplo simples utilizando o crosstab:

import pandas as pd

# DataFrame exemplo

df = pd.DataFrame({'produto': ['Régua', 'Lápis', 'Régua', 'Lápis', 'Régua', 'Lápis', 'Régua'],

'cliente': ['B2B', 'B2C', 'B2B', 'B2B', 'B2C', 'B2B', 'B2C'],

'quantidade': [100, 5, 50, 200, 10, 50, 3]})

# Aplicando o crosstab

vendas = pd.crosstab(index = df.produto, columns = df.cliente,

values = df.quantidade, aggfunc = "sum")

# Lendo a crosstab

vendas

Neste exemplo, temos um DataFrame com três colunas: produto, cliente, quantidade. Utilizamos a tabela cruzada para agrupar os dados de produto e cliente somando a quantidade de produtos correspondentes a cada um. A tabela abaixo mostra o resultado dessa combinação:

| **cliente** | **B2B** | **B2C** |
| --- | --- | --- |
| produto |  |  |
| Lápis | 250 | 5 |
| Régua | 150 | 13 |

Dessa forma, a função crosstab nos ajuda a obter uma visão clara da distribuição dos dados, a identificar padrões e a entender a relação entre variáveis categóricas.