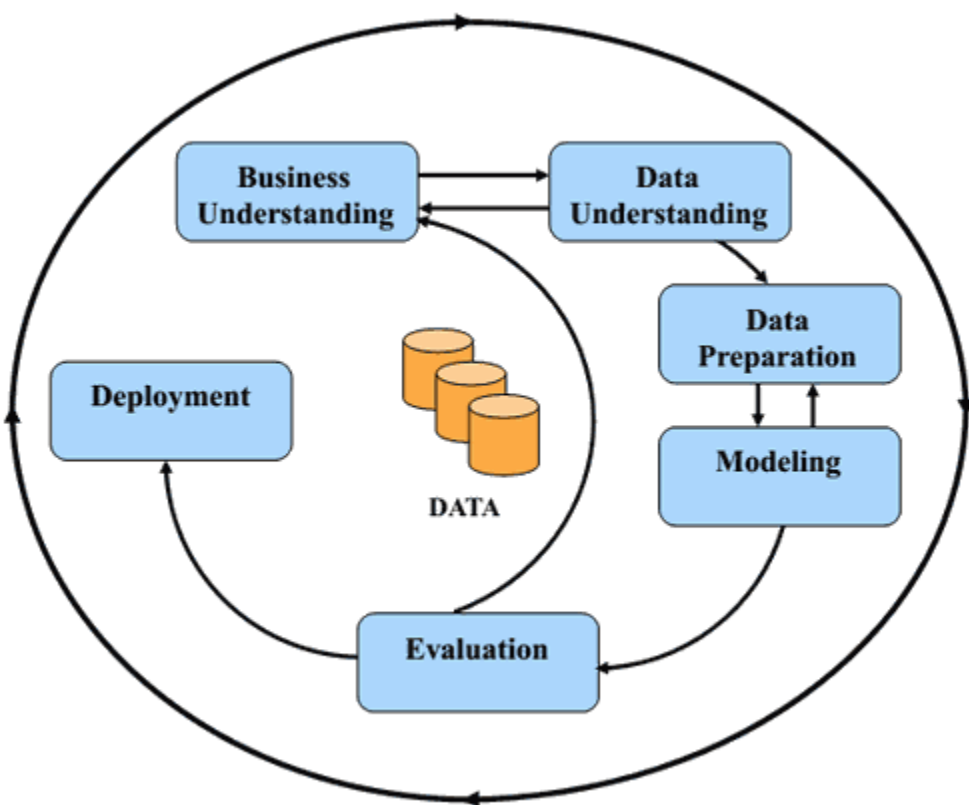
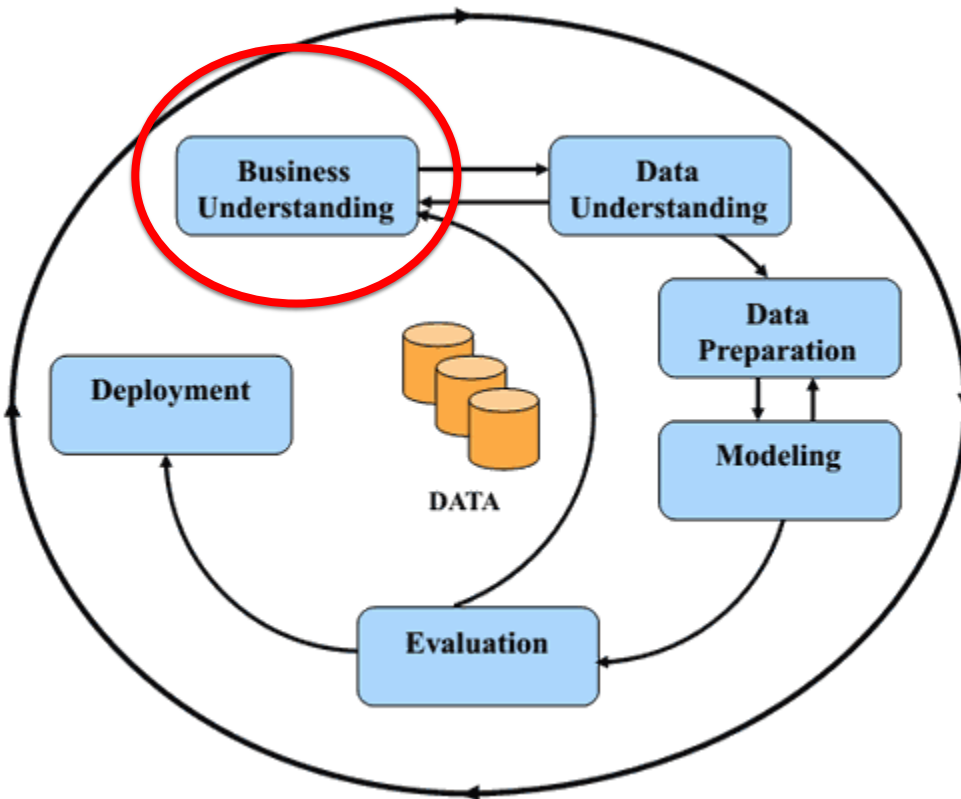


Aula 06



CRISP-DM é uma abordagem estruturada de mineração de dados proposta por um grupo de estudos patrocinado pela união europeia – Cross Industry Standard Process for Data Mining. O modelo abstrai seis passos comuns em projetos de mineração, como na imagem abaixo.



Business Understanding

Principais Objetivos

Definir os critérios de sucesso
Formas de produção?

Como integrar o output com as
tecnologias existentes?

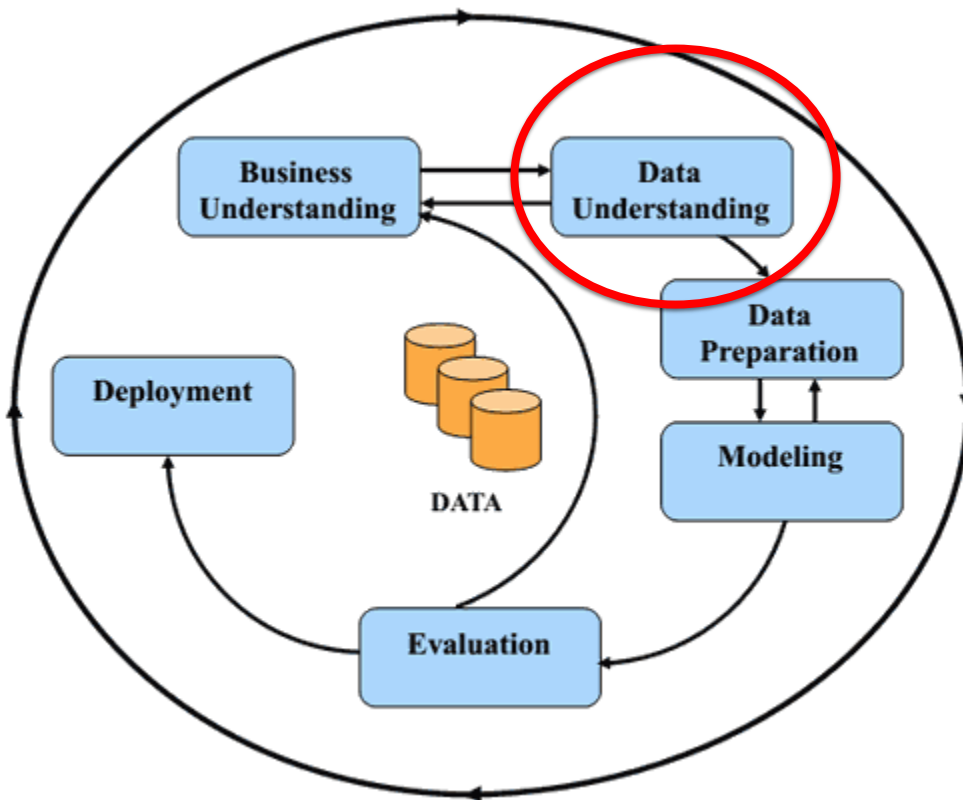
Data Understanding

Principais Objetivos

Recolha de dados

Quais são as fontes de dados?

Análise exploratória de dados
construção de gráficos de dados
simples (histogramas, etc)
para ajudar a compreender a
distribuição de dados



Data Preparation

Principais Objetivos

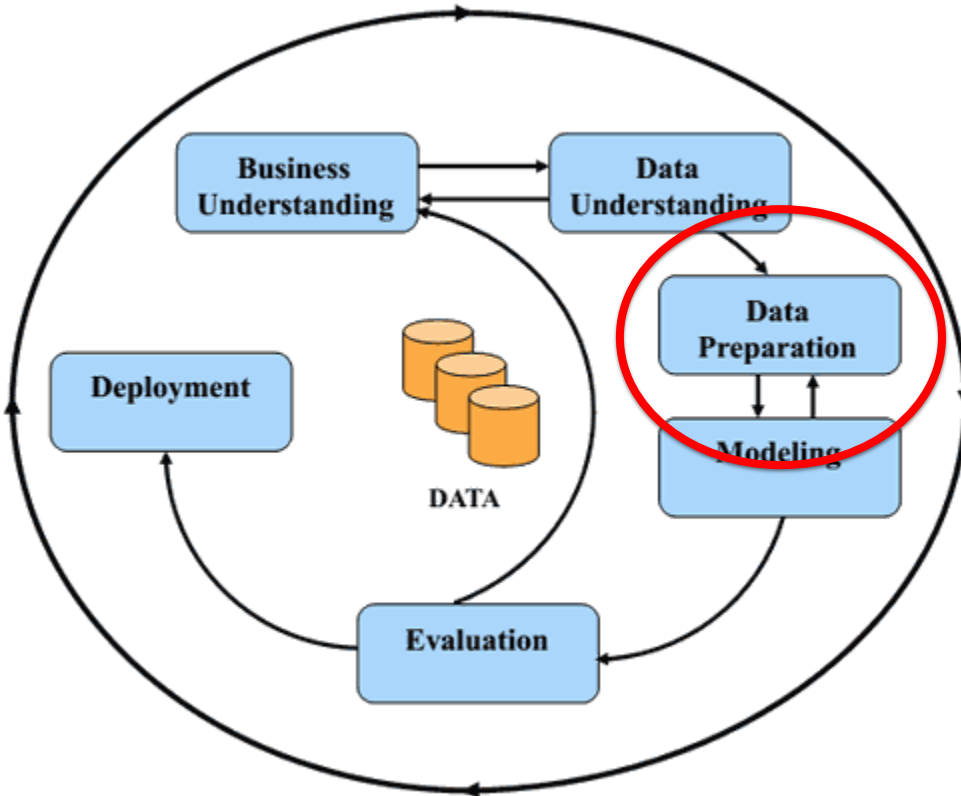
Preparação de dados

Limpeza de Dados - Tratamento de Ruído, Detecção de Anomalias,

Transformação de dados

Normalização de dados,
discretização de dados

Redução de dados



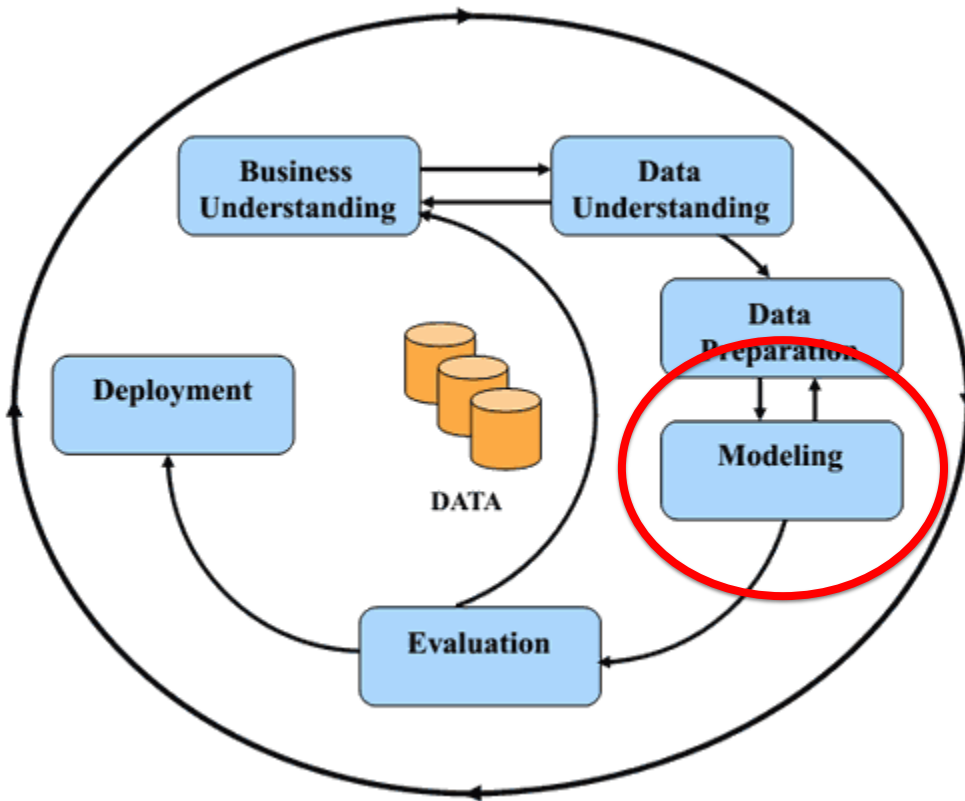
Modeling

Principais Objetivos

Preditiva vs Prescritiva

Preditiva: Modelos de previsão (regressão), classificação

Prescritiva: Achar padrões, entender os clusters, criar insights



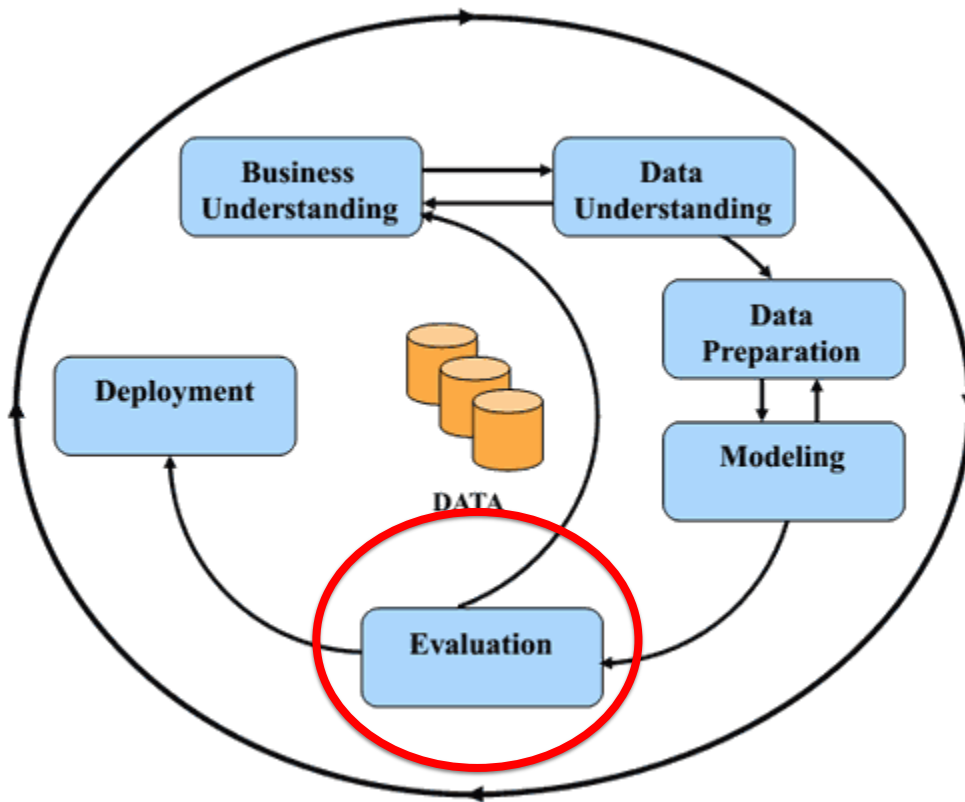
Evaluation

Principais Objetivos

Como avaliar?

Análise dos modelos (cross-validation, error analysis etc)

Análise Subjetiva – análise gráfica, dashboard



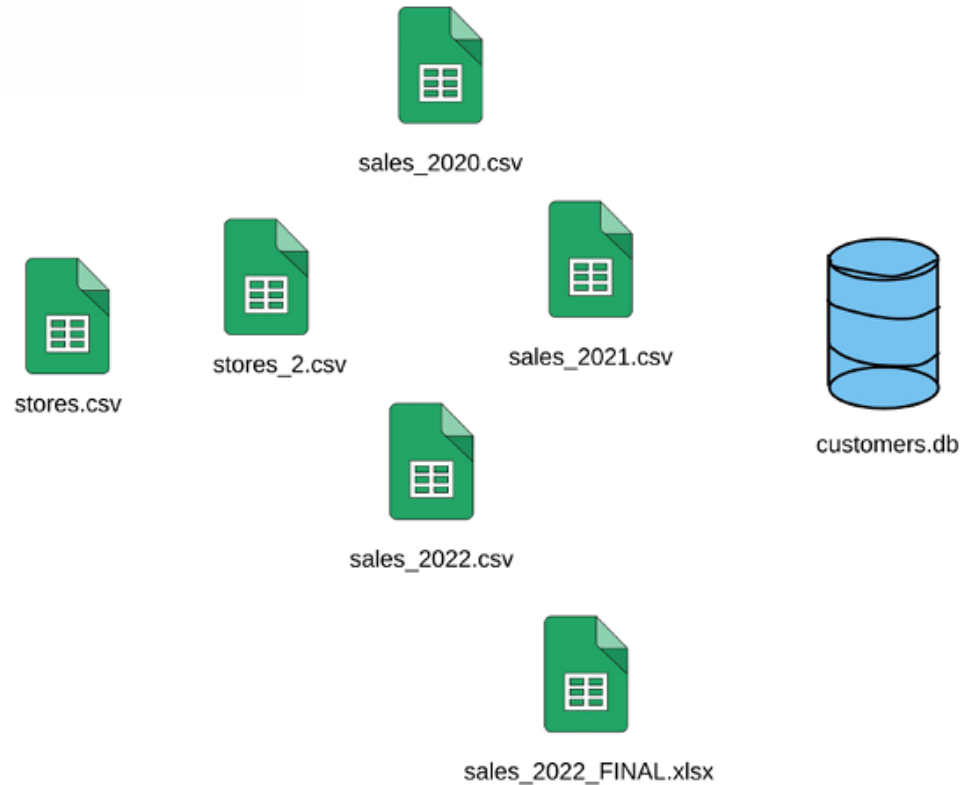
Voltando para Python...

Combinar dados – o ideal



perfect_clean_excel_file_that_answers_all_my_problems.xlsx

Combinar dados – a realidade

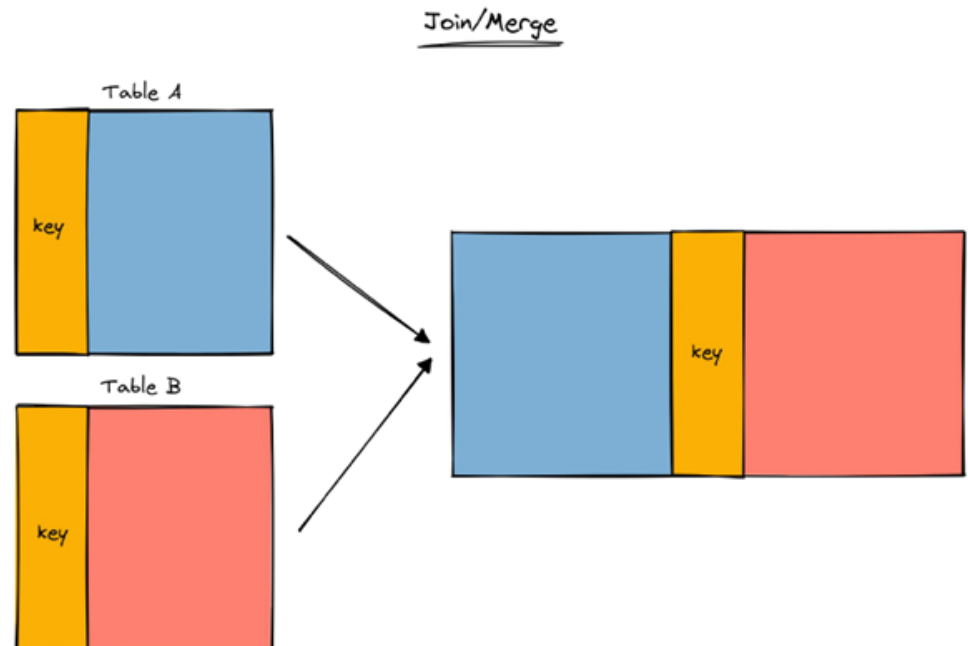


Combinar dados – Join/Merge

Objetivo: ligar observações de 2 ou mais tabelas através de 1 ou mais chaves (keys).

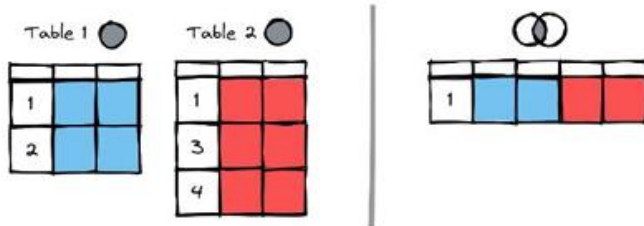
3 questões fundamentais:

- O quê?
- Como?
- Através do quê?

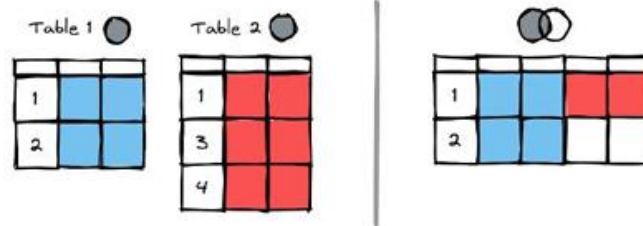


Combinar dados – Operações Join/Merge

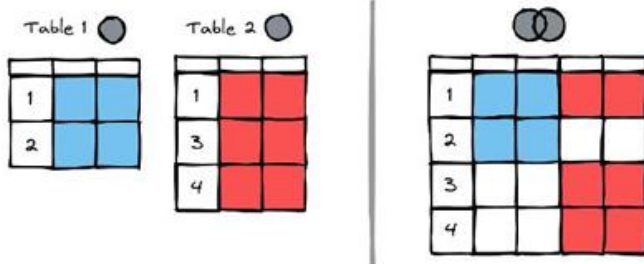
Inner Join



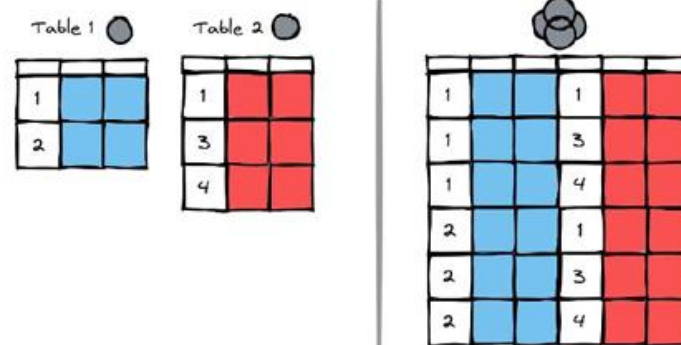
Left Join



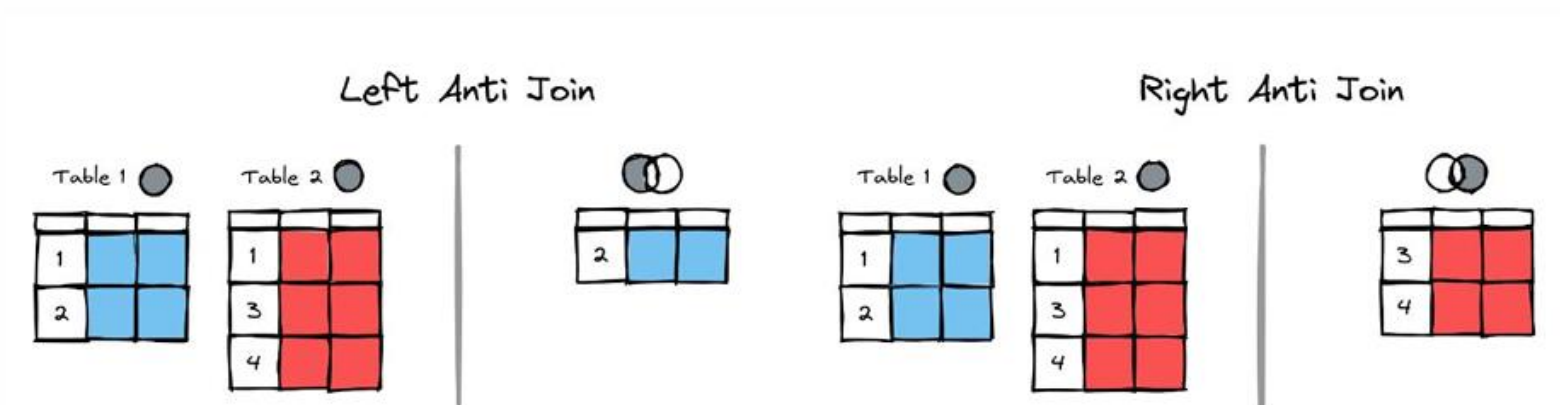
Outer Join



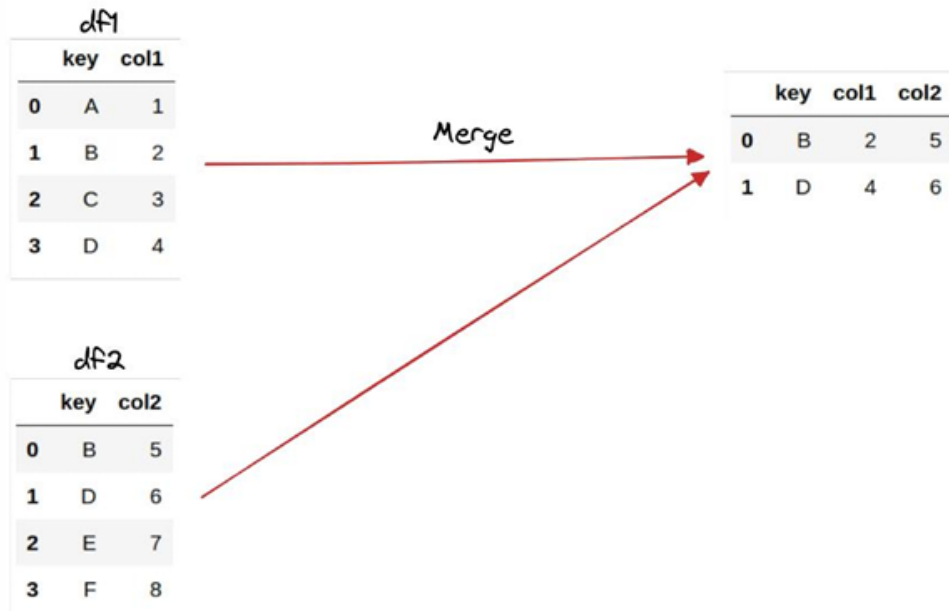
Cross Join



Combinar dados – Anti Joins

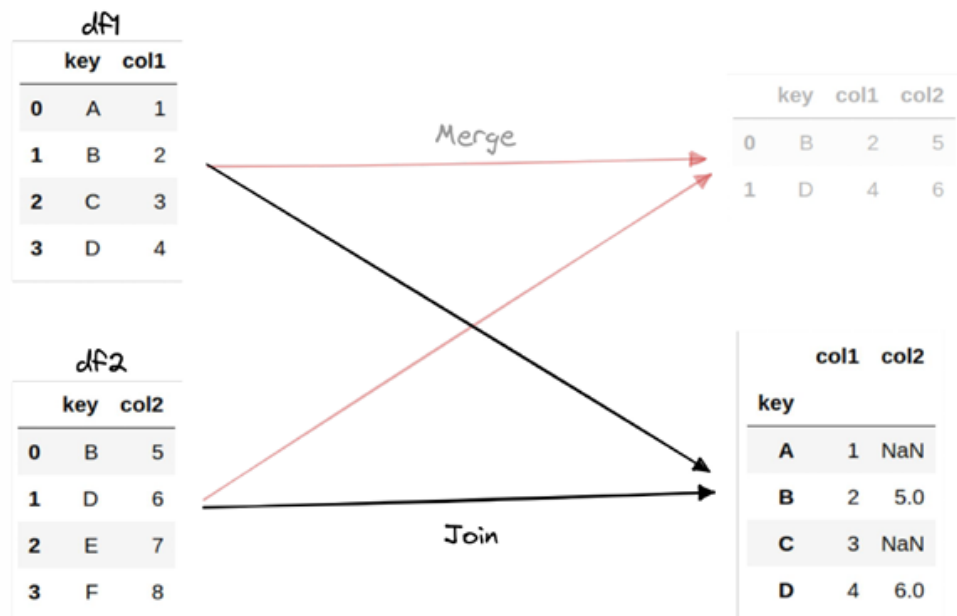


Combinar dados – Join ou Megre?



- Combina os dados com base nos valores de uma coluna
- Diferentes opções de join
- Oferece mais flexibilidade e opções
- Join por default: inner join
- Mais frequentemente usado

Combinar dados – Join ou Merge?



- Combina os dados com base nos valores de uma coluna
- Diferentes opções de join
- Oferece mais flexibilidade e opções
- Join por default: inner join
- Mais frequentemente usado
- Combina os dados com base nos valores dos índices
- Caso não exista, define automaticamente o índice do dataframe final
- Join por default: left join
- Se os dataframes tiverem o mesmo índice, pode ser mais fácil de usar

Combinar dados – Join/Merge

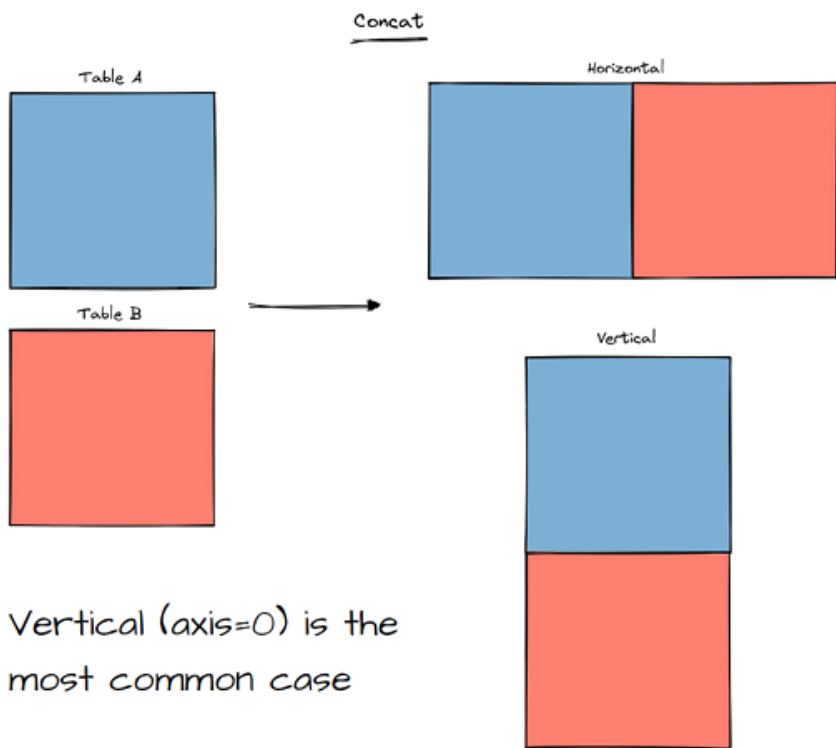
Objetivo: fazer stack dos dados ao longo do eixo (horizontal ou vertical).

Eixo Vertical (axis = 0)

- Se 2 colunas tiverem o mesmo nome, são stacked.
- Se não, são preenchidas com valores nulos.

Eixo horizontal (axis = 1)

- Se 2 colunas tiverem o mesmo nome, são duplicadas.



Vamos praticar!

[Pandas Cheat Sheet.pdf](#)