

# GESTÃO E INTELIGÊNCIA DE DADOS

## TREINAMENTO PYTHON PARA ETL

**Head:** Andrey Alexandre

**Responsável Técnico:** Rodrigo Seares

- Configurações
- Exercícios parte 1
- Exercícios parte 2
- Exercícios parte 3 (Merge)



- **Instalação do Interpretador Python**

- Baixe a python 3.9.4 no link: <https://www.python.org/downloads>  
Quem for utilizar o Linux não precisa instalar, o interpretador python já vem instalado.

- **Instalação da IDE PyCharm**

- Baixe a versão Community no link: <https://www.jetbrains.com/pt-br/pycharm/download/#section=windows>



- **Exercícios**

- Crie um projeto no PyCharm com o nome Treinamento\_Python
- Dentro do PyCharm na raiz, crie um diretório com o Datasets
- Copie o arquivo csv disponível no link:  
[https://itix-my.sharepoint.com/:x:/g/personal/andrey\\_basilio\\_itixti\\_com\\_br/EXZg9g41hAVNnewi4zJv09IBi-NzVvq7hPBXbzHLV7UZdw?e=pEDdnC](https://itix-my.sharepoint.com/:x:/g/personal/andrey_basilio_itixti_com_br/EXZg9g41hAVNnewi4zJv09IBi-NzVvq7hPBXbzHLV7UZdw?e=pEDdnC)
- Armazene no diretório Datasets.
- 1 - Leia o CSV Imóveis.csv para um dataframe.
- 2 – Imprima o número de colunas e o número de linhas do conjunto de dados. (Dica: shape() ).
- 3 – Imprima as 10 primeiras linhas do conjunto de dados (Dica: head() ).
- 4 – Imprima os nomes das colunas do conjunto de dados (Dica: columns ).
- 5 – Imprima o datatype dos dados. (Dica dtypes).
- 6 – Converta a coluna “Data” para o tipo datetime.
- 7 – Crie uma coluna chamada Analise e insira o ano de 2021 nela.
- 8 – Imprima na tela somente as colunas (id, data, preço, quartos, banheiros, ano de construção’)
- 9 – Delete a coluna “metragem da casa.1”

- **Exercícios**

- 10 – Altere o tipo da coluna banheiros para int64.
- 11 – Imprima os imóveis com o ano de construção menor que 2000.
- 12 – Altere o nome de data para DT\_REFERENCIA, preço para VL\_PRECO, quartos para QTD\_QUARTOS, banheiros para QTD\_BANHEIROS, ano de construção para DT\_ANO\_CONSTRUCAO, código postal para NU\_CEP, latitude para NU\_LATITUDE, longitude para NU\_LONGITUDE, andares para QTD\_ANDARES, orla para FL\_ORLA e vista para FL\_VISTA.
- 13 – Crie uma coluna DS\_TIPO\_CONDICAO. Nela classifique o imóvel considerando a seguinte regra:
  - Coluna Condição < 2 então DS\_TIPO\_CONDICAO recebe Ruim.
  - Coluna Condição >= 3 e <= 4 então DS\_TIPO\_CONDICAO recebe Regular.
  - Coluna Condição > 5 então DS\_TIPO\_CONDICAO recebe Bom.
- 14 – Imprima na tela o conjunto de dados ordenados pela coluna VL\_PRECO de forma decrescente. Porém só quero ver as colunas Id e VL\_PRECO.
- 15 – Imprima na tela quantos imóveis estão em condição Regular.
- 16 – Imprima o preço médio das casas que possuem condição Regular.
- 17 – Imprima o preço máximo das casas com 3 quartos.
- 18 – Imprima a quantidade de casas em condição “Bom”, que possua mais de 2 quartos e mais de 2 banheiros.

- **Exercícios**
- 18 – Qual é o ID da casa com maior preço e o id da casa com menor preço.
- 19 – Qual o ano de construção do imóvel mais antigo?



- **Exercícios parte 2**

- 20 – De acordo com a condição do prédio, gostaria de saber o número de prédios pela condição, porém quero que seja utilizado o “groupby”
- 21 – Qual é o preço médio dos imóveis por ano de construção ?
- 22 – Qual é o menor número de quartos por ano de construção de imóveis ?
- 23 – Qual é a soma do preço dos imóveis por número de quartos e banheiros? (Obs: cuidado com a ordem)
- 24 – Qual é a mediana do preço dos imóveis por ano ?
- 25 – Classifique os imóveis de acordo com o seu preço.
  - Preço < 321950 = 0
  - Preço >= 321950 e <= 450000 = 1
  - Preço > 450000 e <= 645000 = 2
  - Preço > 645000 = 3
- 26 – Repita o exercício 24, porém agora faça a classificação usando expressão lambda.
- 27 – Qual é a média de preços dos imóveis por NU\_CEP ?. Exporte os dados para um csv com o cep e a média do preço.

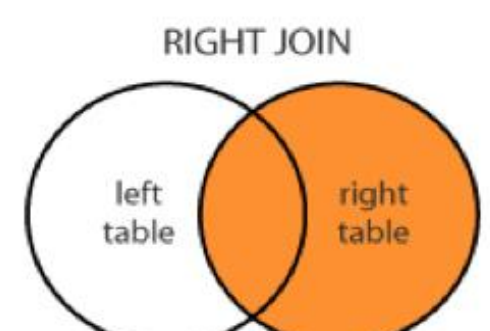
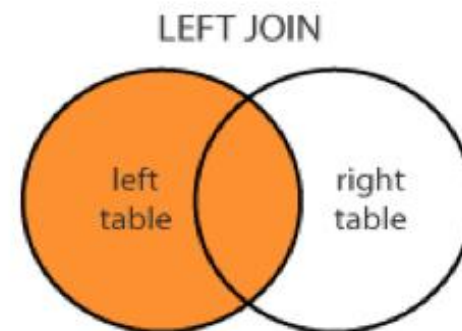
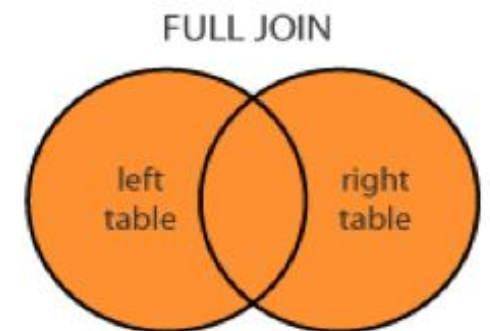
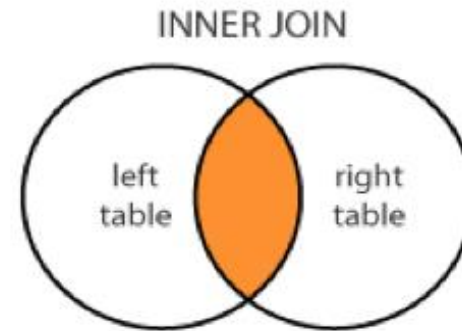
## Tipos de join

**Inner:** Seleciona registros que tenham valores em ambas tabelas.

**Left :** Seleciona registros da primeira tabela (esquerda) com registros da tabela da direita correspondente.

**Right:** Seleciona registros da segunda tabela(direita).

**Full:** Seleciona todos os registros que combinam com os registros da tabela da direita ou da esquerda.





- **Exercícios parte 3 (Merge)**

- 1 – Leia os xlsx (CadastroLoja\_A, CadastroLoja\_B, TipoCliente e Vendas) e jogue para um dataframe respectivamente.
- 2 – Analise os dados (head, shape, dtypes), verifique se é necessário alterar o tipo de dado das colunas.
- 3 - Quais clientes da loja A fizeram compras ?
- 4 - Quais clientes da loja B fizeram compras ?
- 5 - Quais clientes estão cadastrados nas duas lojas ?
- 6 - Percebeu que as colunas ficaram com \_x e da loja b \_y? Quero que troque o sufixo para "\_a" para a loja A e "\_b" para a loja B.
- 7 - Quero ter uma base com todos os clientes de todas as lojas.
- 8 - Se você analisou o resultado, percebeu que existe registro repetidos, remova os registros repetidos.
- 9 - Qual é a soma do gasto de todos os clientes da loja A ?
- 10 - Qual é o nome do tipo de cliente nas vendas, apresente o ID do cliente, idade, cep, nome do tipo cliente e valor, não precisa trazer a coluna Data.
- 11 - Quais clientes da loja A não realizaram compras?
- 12 - Quais clientes da loja B não realizaram compras?

- **Exercícios parte 3 (Merge)**

13 - Quero ter uma base com todos os clientes de todas as lojas que ainda não realizaram compras. Obs: Não quero registros repetidos.

14 - Utilize o seu Data Frame com todos os clientes da loja A e B.(Exercicio 8). Faça um merge com a tabela vendas de forma que só retorne os clientes que já realizaram compras.

15 - Quais são os três clientes que mais compraram? Traga o ID, Nome, e Valor.

