Welcome to

INTERNSHIP STUDIO

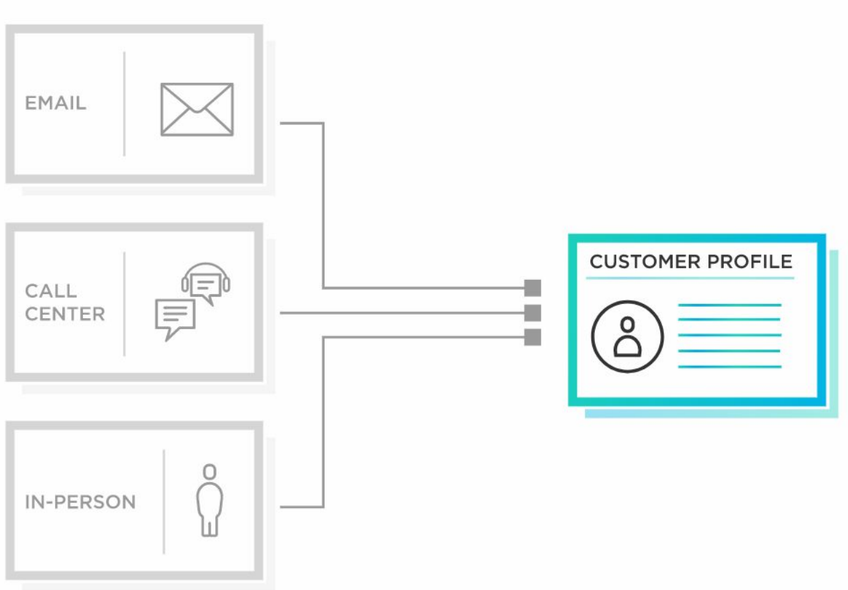
Module 04 | Lesson 04

Data Manipulation

Merging Data with Pandas

Introduction to Data Merging

- Data often comes from multiple sources and needs to be combined for comprehensive analysis.
- Pandas provides powerful tools for merging and joining datasets efficiently.

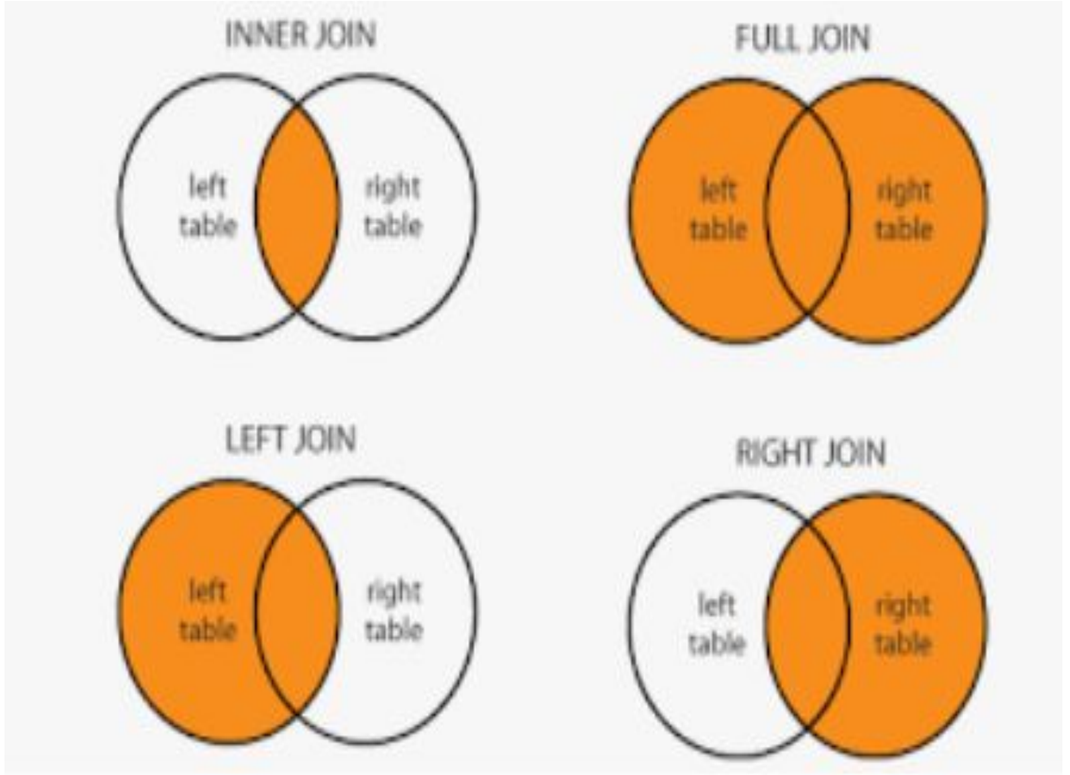


Why Merge data?

- Merge data to gain insights from multiple sources.
- Combine related information for analysis and decision-making.
- Create unified datasets for further processing.

Types of Joins

- **Inner Join**: Returns the intersection of two datasets based on a common key.
- **Left Join**: Returns all records from the left dataset and matching records from the right dataset.
- **Right Join**: Returns all records from the right dataset and matching records from the left dataset.
- **Outer Join**: Returns all records when there is a match in either the left or right dataset.



Merging Two Dataframes

internship studio

- Use the **merge()** function in Pandas to merge two DataFrames.
- Specify the common key using the **on** parameter.
- Choose the appropriate join type using the **how** parameter.

Merging Multiple DataFrames

- Use the **merge()** function iteratively to merge multiple DataFrames.

- Specify the common key and the desired join type for each merge.

internship studio

Handling key Mismatches

- Use the **merge()** function iteratively to merge multiple DataFrames.

- Specify the common key and the desired join type for each merge.

SUMMARY

You got this

- Merging data allows combining information from different sources.
- Pandas provides flexible functions for merging DataFrames.
- Understanding different join types and handling key mismatches is crucial.

Next session

Live coding on Jupyter Notebook