

Introduction to Programming and Numerical Analysis

Exercise Class 7 Exercise 7

Jonas Theodor Ø. Schmidt

UNIVERSITY OF COPENHAGEN



Beware!

- I cannot grade or approve your assignment unless you have specified your full name or/and KU-id in the README.md file for your group on GitHub.
- Remember to provide peer feedback by the 31st of March.

Today's Program

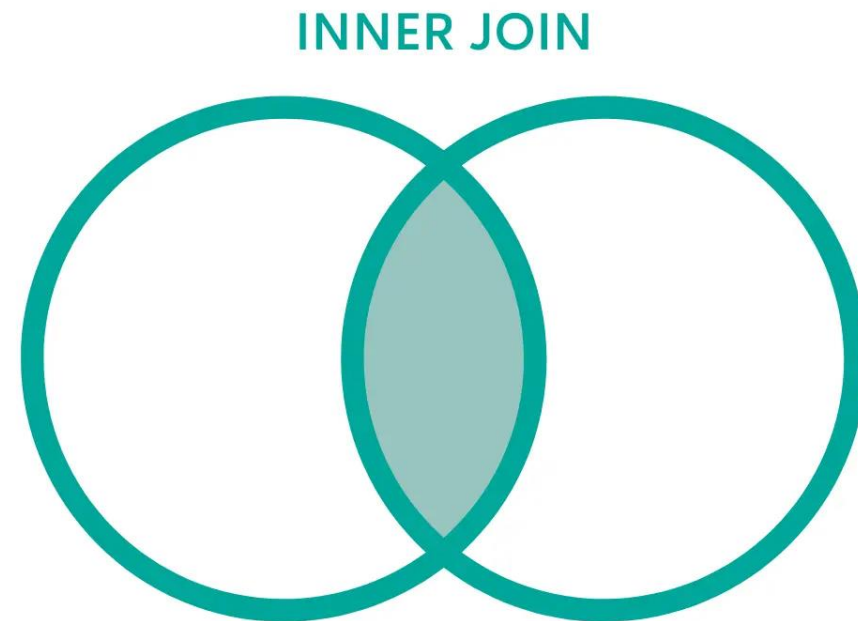
- 15:15 – 15:20: Introduction to Problem Set 4 and Merges
- 15:20 – 15:50: Work on Problem Set 4
- 15:50-16:00: Review of A2 – A4
- 16:00 – 16:15: Break
- 16:15 – 16:50: Work on Problem Set 4
- 16:50-17:00: Review of A5-A6

Introduction to Problem Set 4

- Pandas are essential for data manipulation in Python
- Pandas are syntax heavy and may not seem intuitive at first
- Explore all available resources to find solutions before consulting the answer!
 - Lectures: 'Combining Datasets', 'Data: Fetching data' and 'Data: Split-Apply-Combine'
 - Google, Stack Overflow and ChatGPT
 - [Cheat sheet from DataCamp](#)

Merges – Inner Join

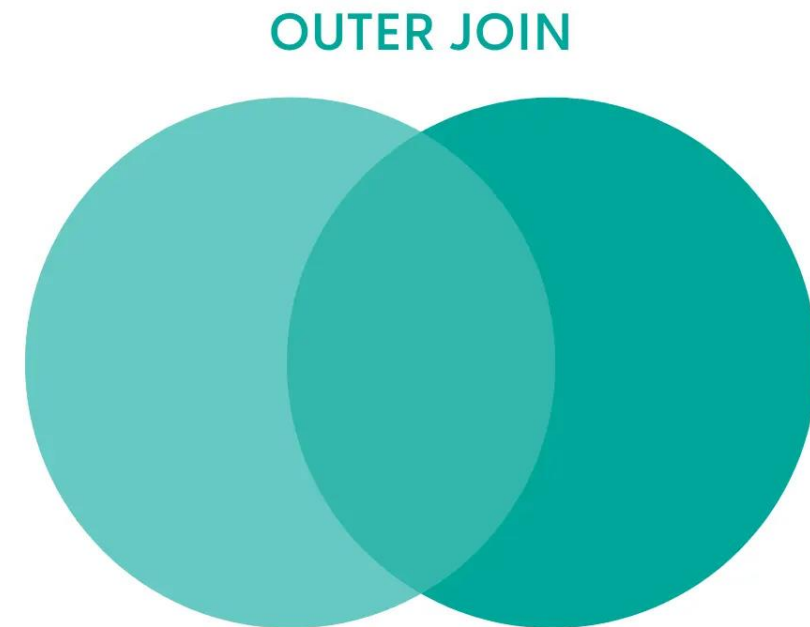
- The inner join in pandas merges two DataFrames at their intersection, returning only the rows common to both.



Source: stratascratch

Merges – Outer Join

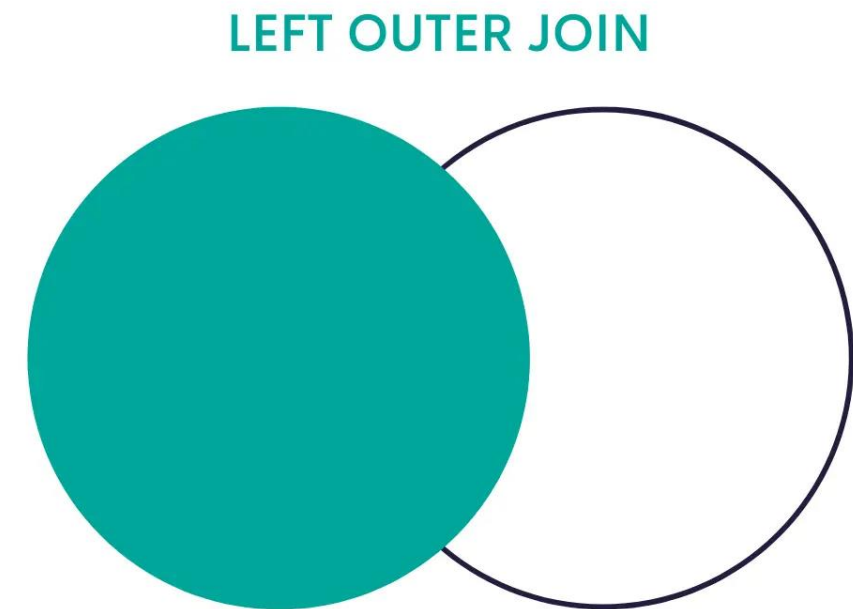
- Outer join in pandas returns all rows from both DataFrames, matching rows that exist in either one.
Rows that is only found in one DataFrame will be filled with NA-values.



Source: stratascratch

Merges – Left Outer Join

- Left outer join in pandas returns all rows from the left DataFrame, merging them with matching rows from the right DataFrame. Unmatched rows from the right DataFrame are filled with NA values.



Source: stratascratch



Questions & comments?