Data Science: Tools & Process
Week 3 (Module 4!)

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Agenda

- **Announcements**
  - Today's Lecture is on Module 4. Next Week's lecture is on Module 5. Module 3 is postponed for 2 weeks.
  - LinkedIn. Please post at least once a week.
  - Milestone Project 1. Redo the DFD
  - Milestone Project 2. Select a dataset from Milestone 2 Overview
- **Last Week's Assignment:**
  - Review DFD Semantics. Do not use the word "data". Strict grading. Milestone is worth 50 points
- **Relational Algebra (Continued from last time)**
- **Break**
- **Aberrant Data Lab**
  - L04-A-1-RemoveOutliers.py
  - L04-A-2-ReplaceOutliers.py
- **Lesson 04 Quiz 1 (Quiz on Aberrant Data)**
- **Data Type Lab**
  - L04-B-1-DataTypes.py
  - L04-B-2-RemoveMissing.py
- **Lesson 04 Quiz 2 (Quiz on Data Types)**
- **Break**
- **Lesson 05 Lab (Time Permitting)**
  - L05-A-1-Dataframes.py, L05-A-2-DataframesCSV.py, L05-B-1-PandasEDA.py

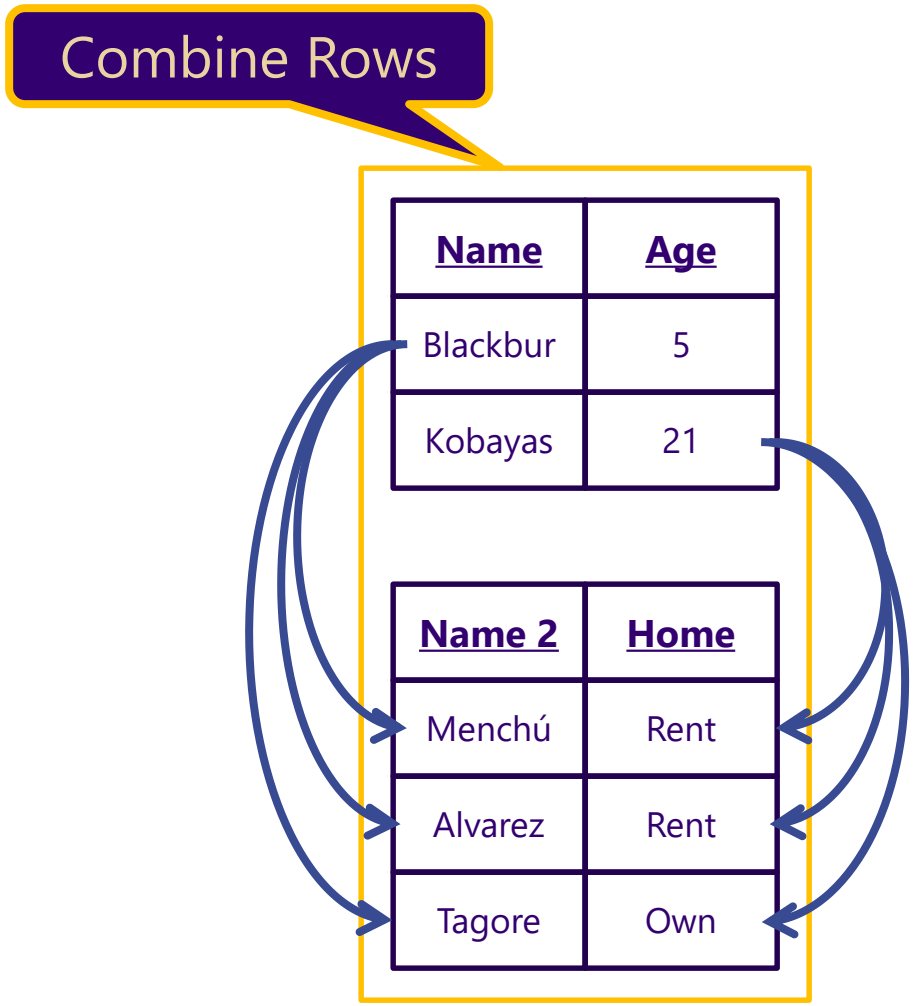
Relational Algebra

The Theory behind Relational Databases

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Product Operation

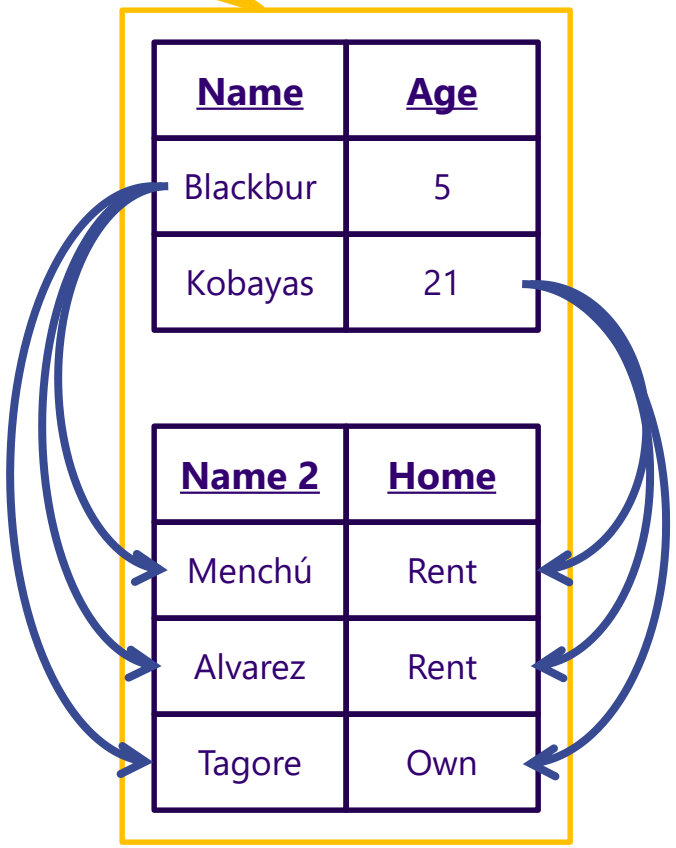
Relational Algebra: Product



Relational Algebra Product:
R X S

Relational Algebra: Product

Combine Rows

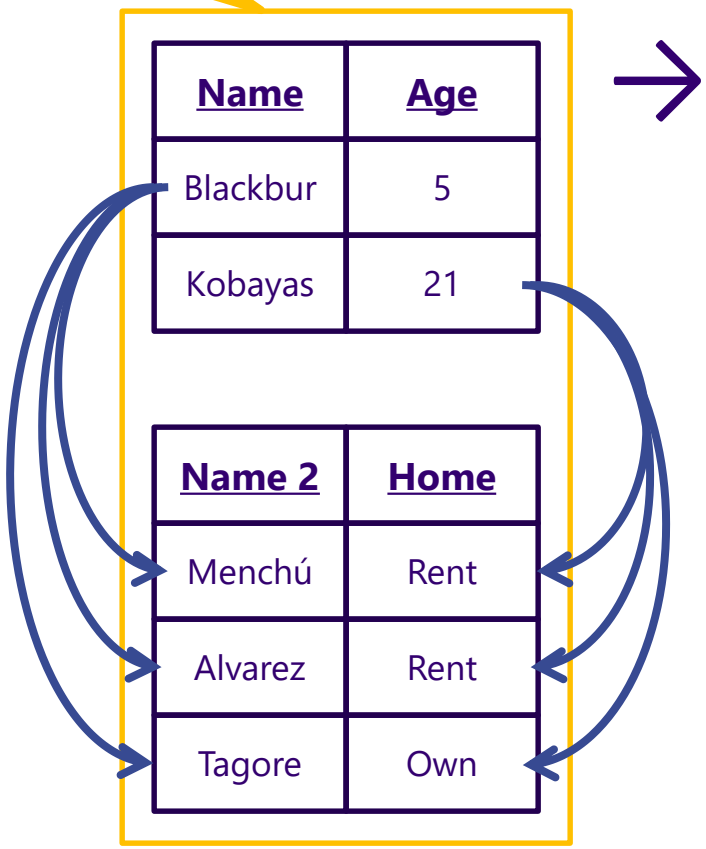


SQL Statement:
SELECT * FROM TableR, TableS

Relational Algebra Product:
R X S

Relational Algebra: Product

Combine Rows



Name	Age
Blackbur	5
Kobayas	21

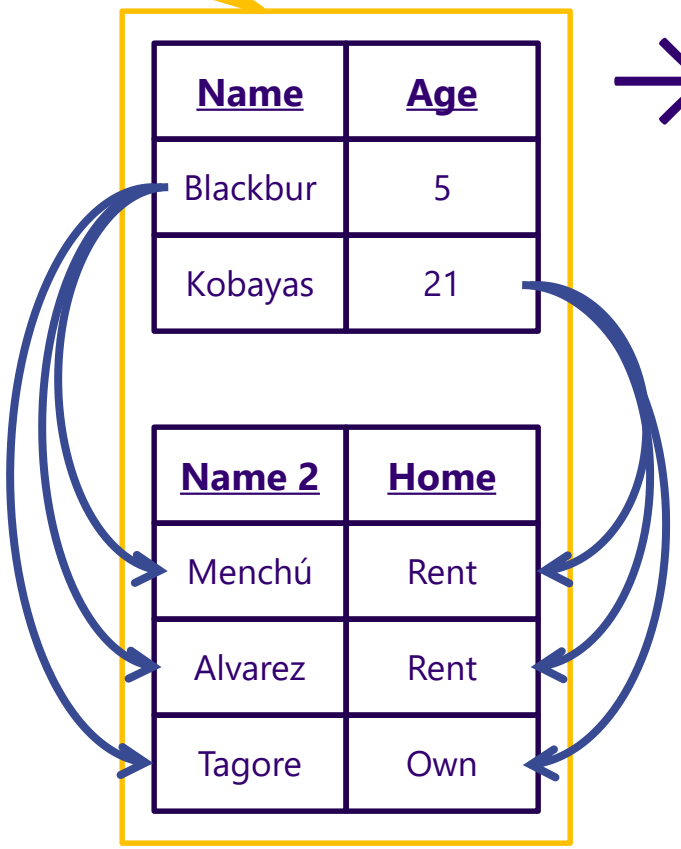
Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
		Alvarez	Rent
		Tagore	Own
Kobayas	21	Menchú	Rent
		Alvarez	Rent
		Tagore	Own

Relational Algebra Product:
R X S

Relational Algebra: Product

Combine Rows



Name	Age
Blackbur	5
Kobayas	21

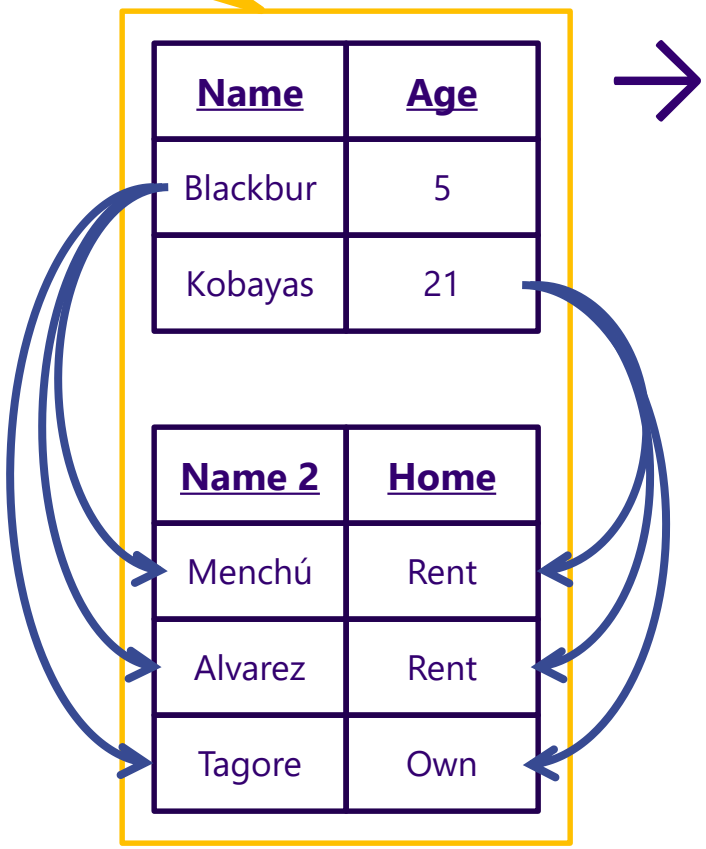
Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own

Relational Algebra Product:
R X S

Relational Algebra: Product

Combine Rows



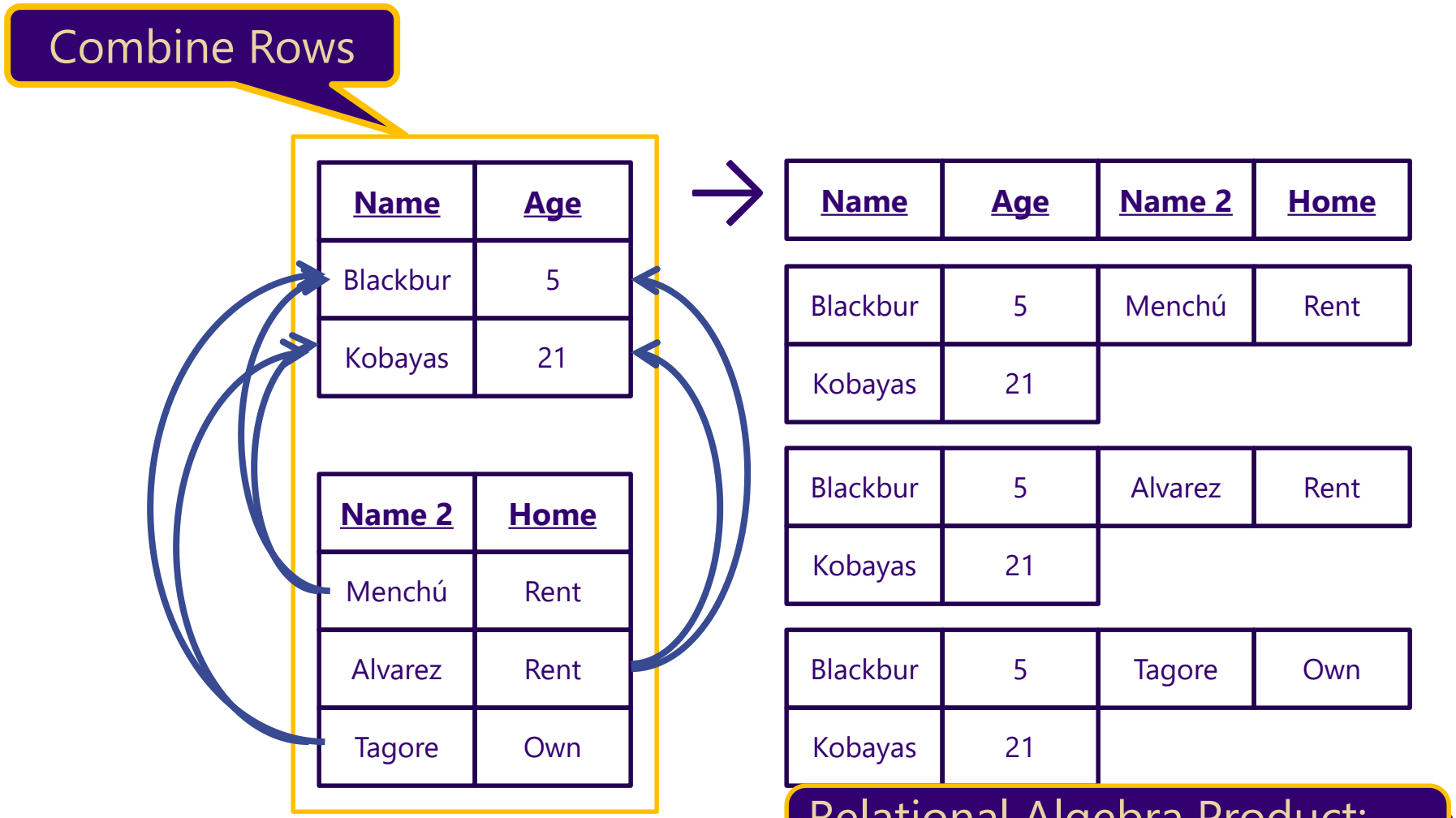
Name	Age
Blackbur	5
Kobayas	21

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own

Relational Algebra Product:
R X S

Relational Algebra: Product



Relational Algebra: Product

Combine Rows

Name	Age
Blackbur	5
Kobayas	21

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

→

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Kobayas	21	Menchú	Rent
Blackbur	5	Alvarez	Rent
Kobayas	21	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Tagore	Own

Relational Algebra Product:
R X S

Relational Algebra: Product

Combine Rows

Name	Age
Blackbur	5
Kobayas	21

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

→

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Kobayas	21	Menchú	Rent
Blackbur	5	Alvarez	Rent
Kobayas	21	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Tagore	Own

Relational Algebra Product:
R X S

Relational Algebra: Product

The result of a product is a relation with n*m tuples where n and m are the number of tuples in the operands. The arity of the result is i + j where i and j are the arities of the operands.

Combine Rows

Name	Age
Blackbur	5
Kobayas	21

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Kobayas	21	Menchú	Rent
Blackbur	5	Alvarez	Rent
Kobayas	21	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Tagore	Own

Relational Algebra Product:
R X S

Relational Algebra: Product

Combine Rows

The result of a product is a relation with n*m tuples where n and m are the number of tuples in the operands. The arity of the result is i + j where i and j are the arities of the operands.

Name	Age
Blackbur	5
Kobayas	21

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own

Relational Algebra Product:
R X S

Join Operation

Relational Algebra: Join

- A Join is a Product with a select statement

- Product followed by Select
  - SELECT * FROM TableR, TableS WHERE Home = “Rent”
  - σ_{ϕ}(R X S) where ϕ: Home = “Rent”

- JOIN
  - SELECT * FROM TableR JOIN TableS ON Home = “Rent”
  - R $\bowtie_{\phi}$ S where ϕ: Home = “Rent”

Relational Algebra: Join

Combine Rows

Name	Age
Blackbur	5
Kobayas	21

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

→

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Kobayas	21	Menchú	Rent
Blackbur	5	Alvarez	Rent
Kobayas	21	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Tagore	Own

Relational Algebra Product with Select:

$\sigma_{\phi}(R \times S)$ where ϕ: Home = "Rent"

Relational Algebra Join:

$R \bowtie_{\phi} S$ where ϕ: Home = "Rent"

Relational Algebra: Join

Combine Rows

Name	Age
Blackbur	5
Kobayas	21

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Kobayas	21	Menchú	Rent
Blackbur	5	Alvarez	Rent
Kobayas	21	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Tagore	Own

Relational Algebra Product with Select:
$\sigma_{\phi}(R \times S)$ where ϕ: Home = "Rent"
Relational Algebra Join:
$R \bowtie_{\phi} S$ where ϕ: Home = "Rent"

Relational Algebra: Join

Combine Rows

Name	Age
Blackbur	5
Kobayas	21

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Kobayas	21	Menchú	Rent
Blackbur	5	Alvarez	Rent
Kobayas	21	Alvarez	Rent

Relational Algebra Product with Select:
$\sigma_{\phi}(R \times S)$ where ϕ: Home = "Rent"
Relational Algebra Join:
$R \bowtie_{\phi} S$ where ϕ: Home = "Rent"

Relational Algebra Operations

So far:

- Union
- Intersect
- Project
- Select
- Product
- Join

Coming up:

- Division

Division Operation

Relational Algebra: Division

This was a Product Operand

Name	Age
Blackbur	5
Kobayas	21

This was the result of a Product

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

This was a Product Operand

Relational Algebra Division:
R ÷ S

Relational Algebra: Division

A Division is sort of like the reverse of a Product

This was a Product Operand

This was the result of a Product

Name	Age
Blackbur	5
Kobayas	21

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

This was a Product Operand

Relational Algebra Division:
R ÷ S

Relational Algebra: Division

A Division is sort of like the reverse of a Product

This was a Product Operand

Name	Age
Blackbur	5
Kobayas	21

This was the result of a Product

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

This was a Product Operand

Relational Algebra Division:
R ÷ S

Relational Algebra: Division

Name	Age
Blackbur	5
Kobayas	21

↓

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

←

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own
Sancar	54	Tagore	Own

Add another row to this table that did not result from the product.

Relational Algebra Division:
R ÷ S

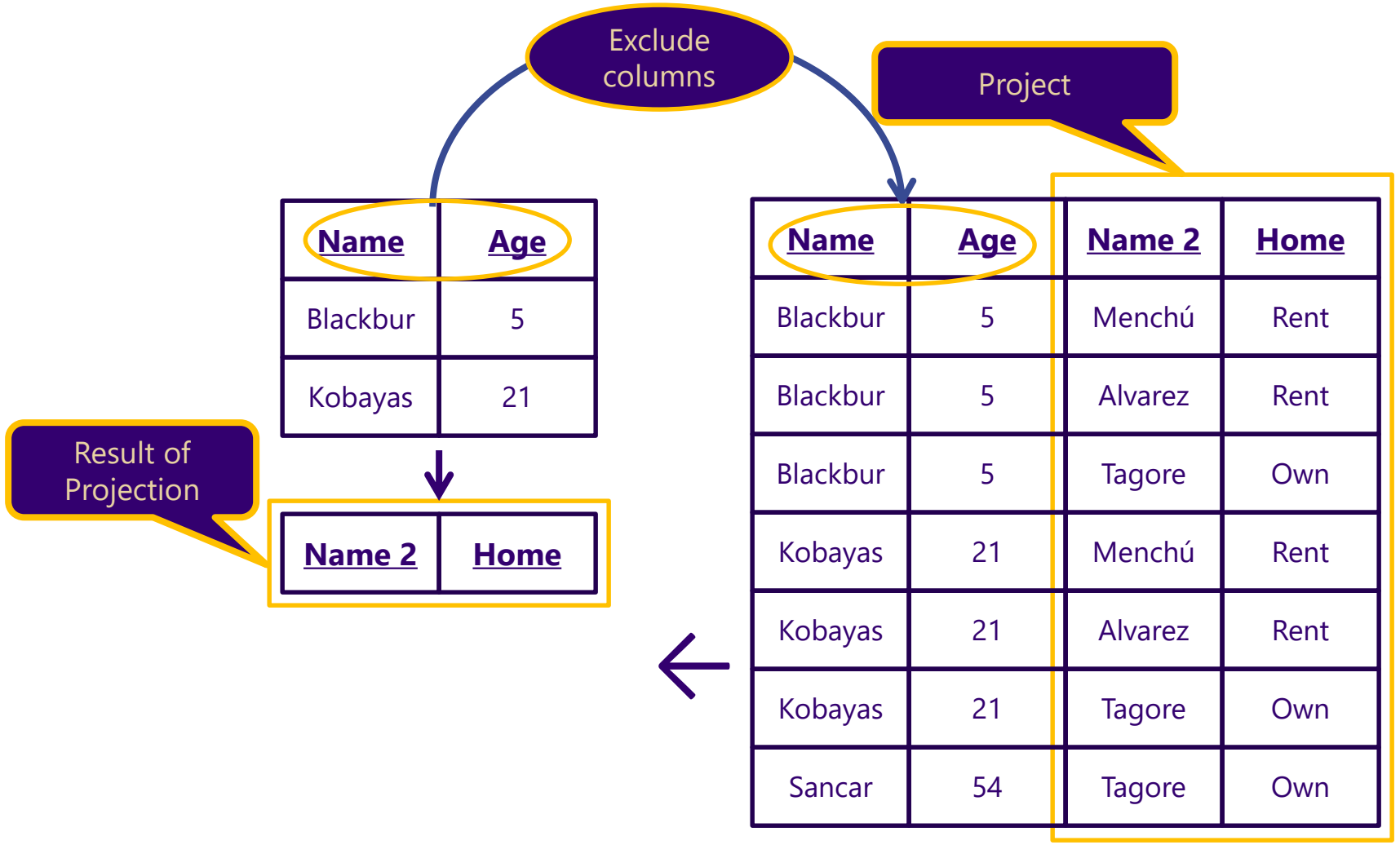
Relational Algebra: Division

Name	Age
Blackbur	5
Kobayas	21

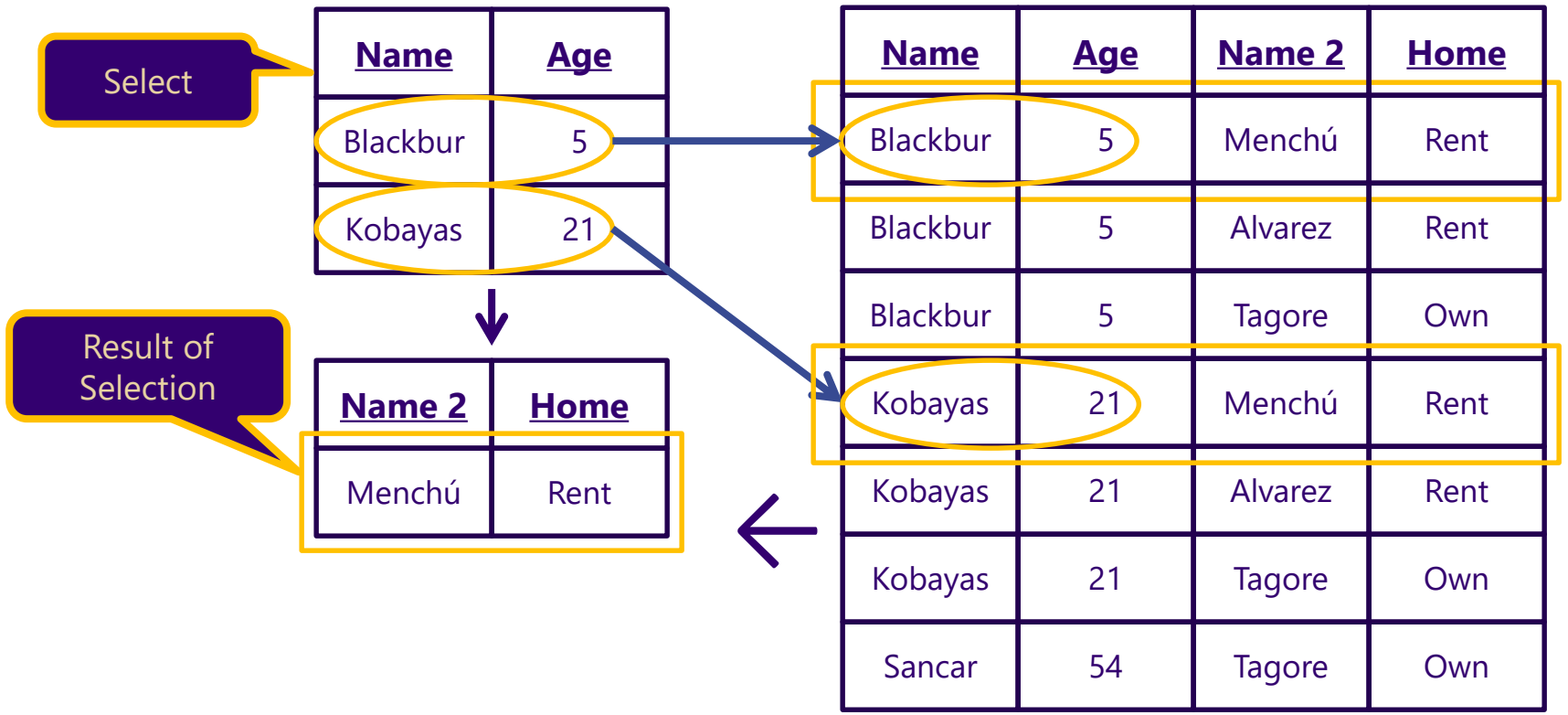
Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own
Sancar	54	Tagore	Own

Relational Algebra Division:
R ÷ S

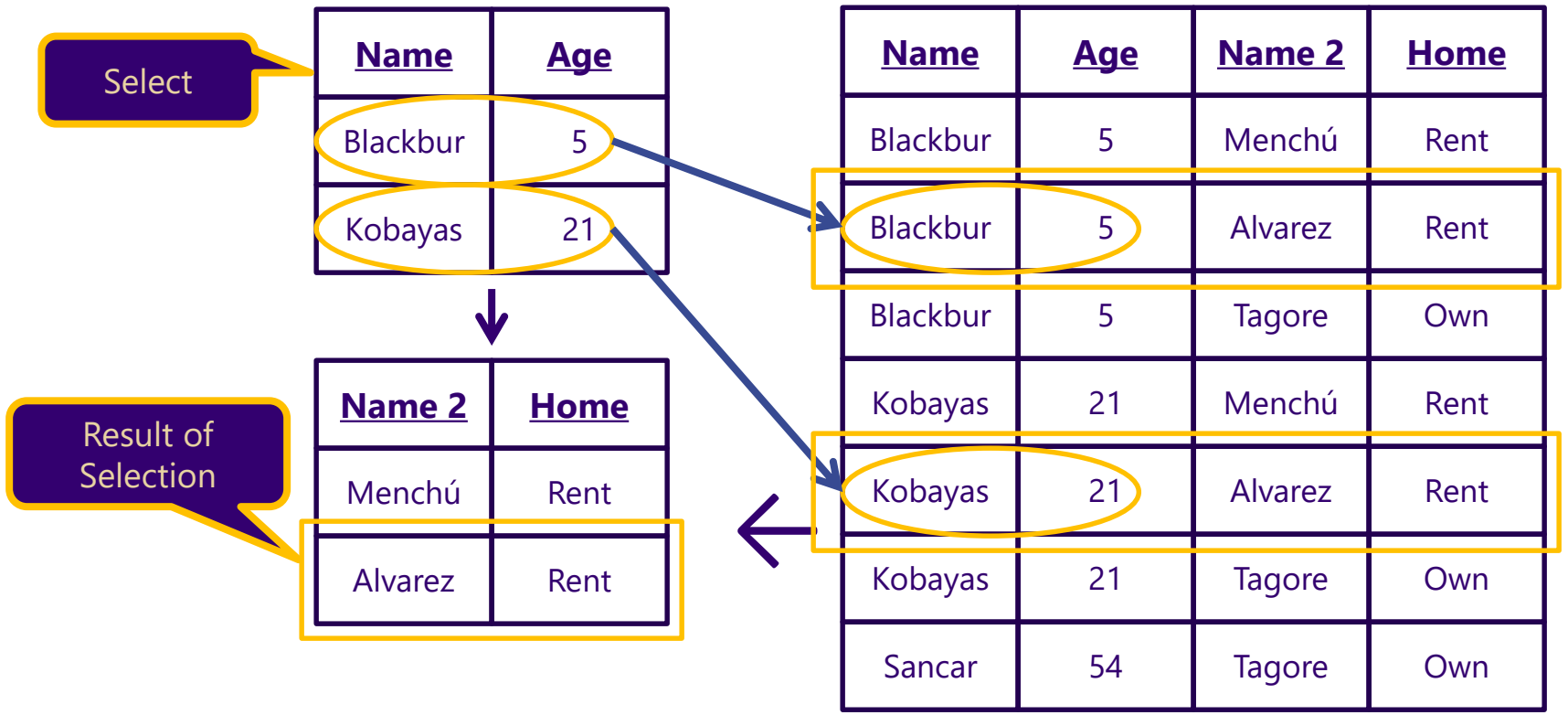
Relational Algebra: Division



Relational Algebra: Division

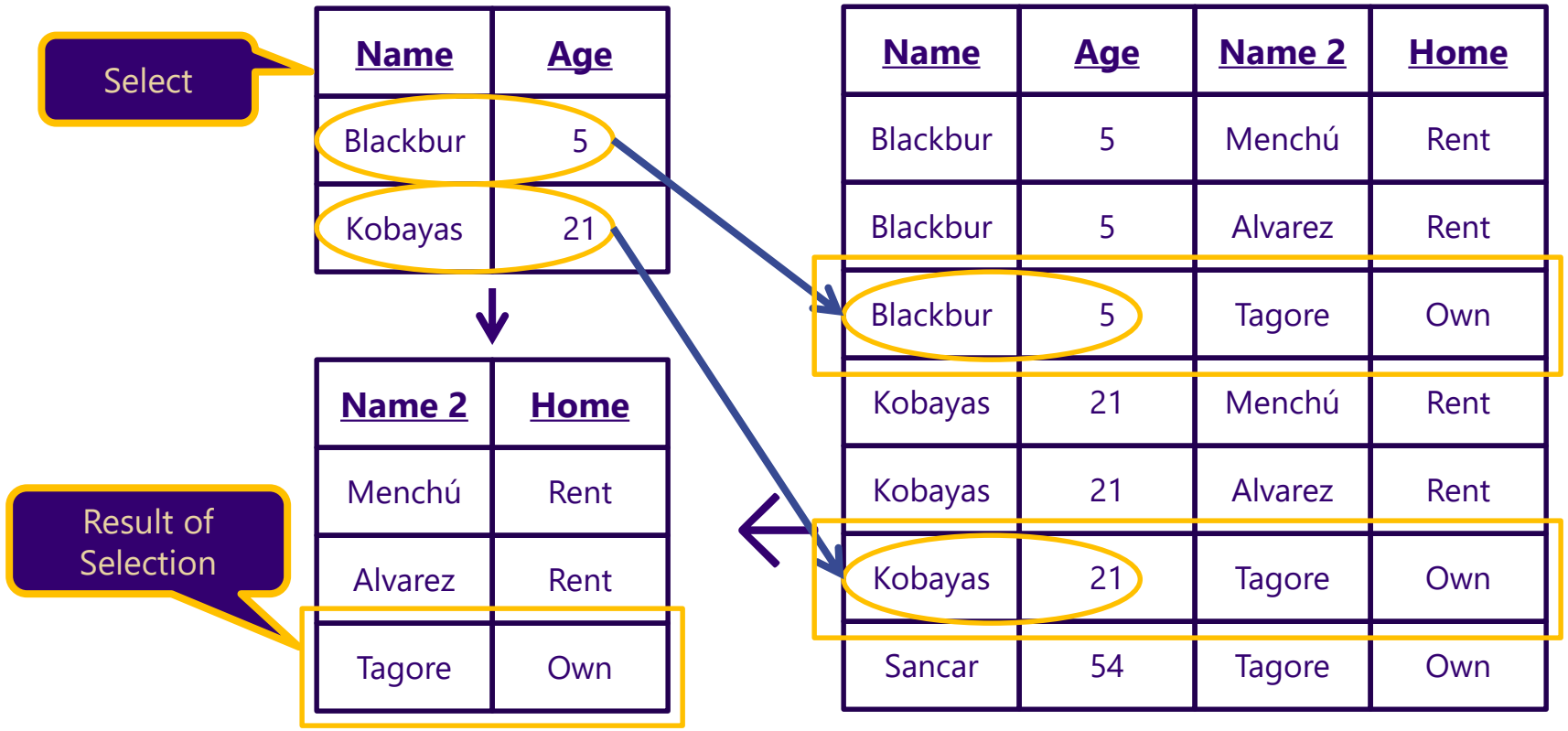


Relational Algebra: Division



Relational Algebra Division:
R ÷ S

Relational Algebra: Division



Relational Algebra: Division

[Menchú, Rent] is in the same tuple as [Blackbur, 5] and [Kobayas, 21]

Name	Age
Blackbur	5
Kobayas	21

↓

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

←

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own
Sancar	54	Tagore	Own

Relational Algebra Division:
R ÷ S

Relational Algebra: Division

[Alvarez, Rent] is in the same tuple as [Blackbur, 5] and [Kobayas, 21]

Name	Age
Blackbur	5
Kobayas	21

↓

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

←

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own
Sancar	54	Tagore	Own

Relational Algebra Division:
R ÷ S

Relational Algebra: Division

[Tagore, Own] is in the same tuple as [Blackbur, 5] and [Kobayas, 21]

Name	Age
Blackbur	5
Kobayas	21

↓

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

←

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own
Sancar	54	Tagore	Own

Relational Algebra Division:
R ÷ S

Relational Algebra: Division

The result of a division is a relation with n tuples of arity I where the dividend operand has at least n*m tuples of arity i + j and the divisor operand has exactly m tuples of arity j that are a subset of the of the dividend tuples.

Name	Age
Blackbur	5
Kobayas	21

↓

Name 2	Home
Menchú	Rent
Alvarez	Rent
Tagore	Own

←

Name	Age	Name 2	Home
Blackbur	5	Menchú	Rent
Blackbur	5	Alvarez	Rent
Blackbur	5	Tagore	Own
Kobayas	21	Menchú	Rent
Kobayas	21	Alvarez	Rent
Kobayas	21	Tagore	Own
Sancar	54	Tagore	Own

Relational Algebra Division:
R ÷ S

Relational Algebra: Resources

Links for definitions and concepts:

- http://en.wikipedia.org/wiki/Cartesian_product
- http://en.wikipedia.org/wiki/Commutative_property
- http://en.wikipedia.org/wiki/Associative_property
- http://en.wikipedia.org/wiki/Closure_(mathematics)
- http://en.wikipedia.org/wiki/Relational_calculus
- http://en.wikipedia.org/wiki/Relational_algebra
- http://en.wikipedia.org/wiki/Edgar_F._Codd
- http://en.wikipedia.org/wiki/Relational_model
- http://en.wikipedia.org/wiki/Relational_database
- http://en.wikipedia.org/wiki/Query_language

Summary

- Table = Part of a Database
- Relation = Table with unique rows
- Attribute = Column in a table relation
  - Arity – number of columns
- Tuple = Row in the table relation
- Math operations on a Relation
  - Union, Intersect, Project, Select, Join
  - Product, Division

Relational Algebra

The theory behind Relational Databases

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Break

Break

Aberrant Data Lab

- Files for Lab:
  - L04-A-1-RemoveOutliers.py
  - L04-A-2-ReplaceOutliers.py

Lesson 04 Quiz 1

Quiz on Aberrant Data

Answer: float(“nan”)

Remove Missing Values Lab

- Files for Lab:
  - L04-B-1-DataTypes.py (We need to understand data types in the context of missing values)
  - L04-B-2-RemoveMissing.py

Lesson 04 Quiz 2

Quiz on Data Type

Break

Break