

COMP 543: Tools & Models for Data Science

Lecture 3 Relational Calculus Handout

FREQUENTS (DRINKER, CAFE)

SERVES (CAFE, COFFEE)

FREQUENTS

DRINKER	CAFE
Chris	A
Chris	B
Chris	C
Risa	A
Risa	B

SERVES

CAFE	COFFEE
A	Cold Brew
A	Drip
A	Espresso
B	Espresso
C	Drip

We want to know:

Who has not gone to a cafe serving Cold Brew?

Earlier, we established that we can answer the question:

Who has gone to a cafe that serves 'Cold Brew'?

with

$$\{f.DRINKER \mid \text{FREQUENTS}(f) \wedge \exists(s, l)(\text{SERVES}(s) \wedge \\ \wedge s.CAFE = f.CAFE \\ \wedge s.COFFEE = \text{'Cold Brew'})\}$$

That might lead us to think that we could answer:

Who has not gone to a cafe serving Cold Brew?

With

$$\{f.DRINKER \mid \text{FREQUENTS}(f) \wedge \neg \exists(s, l)(\text{SERVES}(s) \wedge \\ \wedge s.CAFE = f.CAFE \\ \wedge s.COFFEE = \text{'Cold Brew'})\}$$

However,

$$\{f.DRINKER \mid \text{FREQUENTS}(f) \wedge \neg \exists(s, l)(\text{SERVES}(s) \wedge \\ \wedge s.CAFE = f.CAFE \\ \wedge s.COFFEE = \text{'Cold Brew'})\}$$

Actually returns: Who has gone to a cafe that does not serve ‘Cold Brew’?

Let’s walk through this expression.

Consider only the data for Risa

1. Do the cross product between FREQUENTS and SERVES
2. Restrict ourselves to the case where s.CAFE = f.CAFE
3. Does the statement evaluate to TRUE?
 - If Yes, then include f.DRINKER in the result set

DRINKER	CAFE	CAFE	COFFEE	Result
Risa	B	B	Drip	T
Risa	A	A	Cold Brew	T
Risa	A	A	Drip	F
Risa	A	A	Espresso	F

- When f.CAFE = ‘B’, Risa gets included in the result set.
- However, based on the information when f.CAFE = ‘A’, Risa should NOT be in the result set
- Issue: We want to look at all the relevant rows together

Who has not gone to a cafe serving Cold Brew?

To answer this question, we need to introduce a second variable:

$$\{f.DRINKER \mid \text{FREQUENTS}(f) \wedge \neg \exists(f_2, s)(\text{FREQ}(f_2) \wedge \text{SERVES}(s) \wedge f_2.CAFE = s.CAFE \wedge s.COFFEE = \text{'Cold Brew'} \wedge f.DRINKER = f_2.DRINKER)\}$$

In this case, by having the second variable, we are able to look at all the data for each Cafe as a whole.

- Here, we have another variable, f_2
- We consider each drinker in turn from the FREQUENTS relation. In this case, look at Risa.

DRINKER
Risa

- Now, look at all the combinations of FREQUENTS and SERVES where the CAFE matches

DRINKER	CAFE	COFFEE	Result
Risa	A	Cold Brew	F
Risa	A	Drip	
Risa	A	Espresso	
Risa	B	Drip	

- If there is any row where the Coffee is 'Cold Brew', we exclude the drinker
- Now, in this case, one of the cafes that Risa frequents does serve Cold Brew, so Risa is not added to the result set