

1 Deanonymisierung (5 Punkte)

Betrachten Sie die folgenden beiden Tabellen. Tabelle 1 zeigt fiktive anonymisierte Daten eines sozialen Netzwerkes, Tabelle 2 soll einer Webseite für Film- und Serienrezensionen entnommen sein. Gehen Sie davon aus, dass alle Personen aus Tabelle 2 auch in Tabelle 1 vorkommen.

Name	Geschlecht	Alter	Stadt	Lieblingsfilm	Lieblingsserie	Beziehungsstatus
*	weiblich	20-23	Neunkirchen	Scream	Rick und Morty	keine Angabe
*	männlich	24-27	Zweibrücken	Star Wars 8	Game of Thrones	in einer Beziehung
*	weiblich	20-23	Marpingen	El Camino	The Big Bang Theory	in einer Beziehung
*	weiblich	24-27	Saarbrücken	El Camino	Moon Knight	keine Angabe
*	männlich	24-27	Zweibrücken	Die Verurteilten	Game of Thrones	Single
*	männlich	20-23	Merzig	Spider-Man 2	Breaking Bad	in einer Beziehung

Tabelle 1: Anonymisierte Daten eines sozialen Netzwerkes.

Name	Film/Serie	Rezensionsdatum	Sternebewertung
David	Game of Thrones	17.08.2020	★★★★★
Dan	Scream	27.11.2019	★★★★★
Lisa	Scream	20.03.2015	★★★★★
Camilla	El Camino	05.04.2022	★★★★★
Dan	Game of Thrones	01.03.2019	★★★★★
Camilla	Spider-Man 2	03.09.2019	★★★★★
Lisa	Breaking Bad	03.02.2016	★★★★★
Tobias	Spider-Man 2	10.05.2018	★★★★★
Tina	The Big Bang Theory	15.03.2017	★
David	Breaking Bad	27.06.2018	★★
Lisa	El Camino	06.07.2020	★
Dan	Spider-Man 2	01.07.2018	★

Tabelle 2: Daten einer Website für Film- und Serienrezensionen.

(a) Welche persönlichen Informationen erhalten Sie über die folgenden Personen durch Verknüpfen der beiden Datenquellen? Erklären Sie Ihr Vorgehen.

1. Tobias
2. David

(b) Ordnen Sie den weiblichen Personen aus Tabelle 1 Namen zu. Begründen Sie Ihr Vorgehen.

2 Krankenkassen und Datenzugriff (5 Punkte)

In dieser Aufgabe beschäftigen wir uns mit einer Krankenkasse und ihren personenbezogenen Daten. Normalerweise speichert sich die Krankenkasse lediglich einige Informationen zu ihren Kunden und Kundinnen, basierend auf folgendem Relationenschema:

[Kunden/Kundinnen] : {[KID: int, Geburtsjahr: int, Monatsbeitrag: float, Beitrittsdatum: date]}

[Behandlungen] : {[BID: int, Beschreibung: text]}

[erhalten] : {[Patient*in:(Kunden/Kundinnen→KID), Behandlung: (Behandlungen→BID),
Datum: date, Kosten: float]}

Durch einen kürzlichen Datenleak hat die Krankenkasse nun allerdings auch (illegalerweise) Zugriff auf externe Datenbanken bekommen, die das Kaufverhalten ihrer Kunden und Kundinnen beschreiben.

[Artikel] : {[AID: int, Name: string, Preis: real, Nährwerte: real]}

[kaufen] : {[Kunde/Kundin:(Kunden/Kundinnen→KID), Artikel:(Artikel→AID), Datum: date,
Menge: float]}

1. Wie kann die Krankenkasse diese Daten, die sie unrechtmäßig erhalten hat, im Zusammenhang mit den bereits vorhandenen Daten missbrauchen, um die Beiträge ihrer Kunden und Kundinnen gewinnbringend anzupassen? (3 Punkte)
2. Formulieren Sie eine oder mehrere umgangssprachliche Anfragen, die potenzielle Kunden und Kundinnen identifizieren, für die die Krankenkasse die Beiträge anpassen sollte. (2 Punkte)

3 Von SQL zum logischen Plan (5 Punkte)

In dieser Aufgabe möchten wir eine SQL Anfrage, die auf dem nachfolgenden Schema basiert, mit den in der Vorlesung vorgestellten Regeln optimieren.

[Personen] : {[PID: int, Name: string, Geburtsjahr: int, Wohnort: string]}

[Schüler*innen] : {[SID:(Personen→PID), Klassenstufe: int, Klassenraum: int, Schulform: string]}

[Klausuren] : {[KID: int, Fach: string, Thema: string, Dauer: int]}

[Lehrer*innen] : {[LID:(Personen→PID), Hauptfach:string, Gehalt:int, Dienstjahre:int]}

[unterrichten] : {[Schüler*in:(Schüler*innen→SID), Datum:date, Uhrzeit:time,
Lehrer*in:(Lehrer*innen→LID), Fach: string]}

[korrigieren] : {[Klausur:(Klausuren→KID), Schüler*in:(Schüler*innen→SID),
Lehrer*in:(Lehrer*innen→LID), Note: int]}

```
SELECT Name, Wohnort, Hauptfach, Gehalt
FROM Personen, Lehrer*innen, unterrichten
WHERE PID = LID
      AND Lehrer*in = LID
      AND Fach = 'Deutsch'
      AND Geburtsjahr < 2000
      AND Gehalt >= 500;
```

- (a) Übersetzen Sie die SQL Anfrage kanonisch in einen Ausdruck der relationalen Algebra. Verwenden Sie dabei nur Projektionen, Selektionen und kartesische Produkte.
- (b) Zeichnen Sie den logischen Plan der Anfrage als Baum.
- (c) Wenden Sie die aus der Vorlesung und dem Notebook Rule-based Optimization.ipynb bekannten Regeln zur heuristischen Anfrageoptimierung an und zeichnen Sie den optimierten, logischen Anfragebaum, der sich dadurch ergibt.

4 Kommissar Equi-Join's schwerster Fall (5 Punkte)

In dieser Aufgabe schlüpfen Sie in die Rolle von Juniorkommissar Gruppierung. Nutzen Sie zum Lösen der Aufgabe das beigelegte Notebook. Der mitgelieferte Datensatz ist eine Erweiterung der Daten des NSA.ipynb Notebooks.

```
[households] : {[id: int, street: string, postcode: int, city: string, floor: int]}
[citizens] : {[id: int, firstname: string, lastname: string, birthday: string]}
[livingIn] : {[citizen_id:(citizens→id), start: string, until: string, household_id:(households→id)]}
[articles] : {[id: int, label: string, unit: int]}
[nutritionalValues] : {[id:(articles→id), calories: int]}
[purchases] : {[article_id:(articles→id), citizen_id:(citizens→id), date: string, amount: real]}
```

Kommissar Equi-Join sprach neulich mit seinem Vorgesetzten und Mentor, Oberkommissar Theta-Join, über alte Fälle, wodurch Equi-Joins Aufmerksamkeit auf einen seiner wenigen, ungelösten Fälle gelenkt wurde. Hierbei geht es um einen Mord an der Person John Doe, welcher sich am 24.11.1943 ereignete. Gemäß dem Autopsiebericht wurde als Todesursache ein sehr seltenes (und fiktives!) Gift festgestellt, das allerdings laut der zuständigen Rechtsmedizinerin, Dr. Selektion, durch eine Liste an Alltagslebensmittel hergestellt werden kann. Diese Liste besteht aus:

- Genau 500 Gramm Gewürzgurken.
- Mindestens zwei Kilogramm Salat.
- Mindestens ein Kilogramm Karotten, aber maximal (inklusive) drei Kilogramm.

Dr. Selektion meinte noch, dass diese Lebensmittel maximal 5 Tage (inklusive) nach dem Einkauf als Gift verwendet werden können, da ansonsten die Wirkung zu schwach wäre.

Basierend auf dieser Annahme, leitete Kommissar Equi-Join damals eine Überprüfung der dokumentierten Einkäufe in den örtlichen Supermärkten ein. Leider konnte er mit diesen Daten nicht viel anfangen. Daher befragte er anschließend Zeugen und Zeuginnen nach verdächtigen Aktivitäten am Tag des Mordes, jedoch blieb auch dies erfolglos.

Kommissar Equi-Join ist nun allerdings davon überzeugt, mit Ihrer Hilfe den Fall lösen zu können. Dazu holt er die alten Befragungen der Zeugen und Zeuginnen heraus. Leider sind die Aussagen aufgrund des Alters der Dokumente zum größten Teil unlesbar geworden. So kann er lediglich eine Seite mit Informationen finden, in denen Zeugen und Zeuginnen davon berichten, wie verdächtige Personen am Tag des Mordfalls in ihre Haushalte zurückkehren. Dabei sind die folgenden Informationen über die Adressen der Verdächtigen noch lesbar:

- Adresse 1: ...18
- Adresse 2: ...straße 1...
- Adresse 3: L...

Kommissar Equi-Join gibt Ihnen in diesem Zusammenhang das damalige Einwohnerregister, in dem Informationen über die örtlichen Bewohner und deren gemeldeten Häuser zu finden sind, sowie die Daten über die damals registrierten Einkäufe. Können sie aufgrund dieser Daten Kommissar Equi-Join helfen, seinen alten Fall zu lösen? Geben Sie ihre Lösung als SQL-Anfrage ab, die die folgende Ausgabe hat:

- Die Vornamen der Verdächtigen als 'Vorname'.
- Die Nachnamen der Verdächtigen als 'Nachname'

Sie dürfen zum Lösen dieser Aufgabe Unteranfragen sowie Views benutzen. Erläutern Sie zudem in der `jupyter.txt`, ob Sie anhand der ausgegebenen Daten eine*n Hauptverdächtige*n eindeutig identifizieren können.

Abgabe

Lösungen sind in Teams von 2 bis 3 Studierenden bis zum 2. Juni 2022, 10:15 Uhr über Ihre persönlichen Statusseite im CMS einzureichen. Nutzen Sie hierfür die Team Groupings Funktionalität im CMS.

Ihre Abgabe muss dem folgenden Format entsprechen:

```
abgabe.zip
├── abgabe.pdf
└── jupyter.txt
```

Hierbei enthält `abgabe.pdf` Ihre Lösungen zu Aufgabe 1, 2 und 3 und `jupyter.txt` Ihre Lösung zu Aufgabe 4. Achten Sie darauf, dass Sie nur die von Ihnen zu ergänzenden Jupyter Zellen so kopieren, dass Einrückung und Formatierung korrekt sind.

Abgaben, die nicht den oben angegebenen Vorgaben entsprechen, führen zu Punktabzug. Einzelabgaben werden nicht mehr korrigiert und mit 0 Punkten bewertet.