Aufgabe 1
Finden Sie zu den folgenden Ausdrücken der SQL möglichst äquivalente Versionen in der relationalen Algebra. Geben Sie jeweils eine kurze Begründung an, falls ein Teilausdruck nicht übersetzt werden kann.

a)
```
WITH
  tmp AS (SELECT a,b,c FROM R)
SELECT t1.a, t1.b, t1.c, t2.a, t2.b, t2.c
  FROM tmp AS t1, tmp AS t2
  WHERE true
  ORDER BY t1.a, t2.b ASC
```

b)
```
SELECT L.a, L.b, R.a
  FROM L INNER JOIN R ON L.a = R.b
  WHERE L.c > 0 AND L.c IS NOT NULL
  GROUP BY L.a,L.b,R.a
  HAVING count(L.c) > 5
```

c)
```
WITH
  foo AS
    (SELECT a,b FROM R WHERE a>b
     UNION ALL
     SELECT a,b FROM R WHERE b>a ),
  bar AS
    (SELECT a,b, count(*) as cnt FROM S
        WHERE c IS NOT NULL
        GROUP BY a,b )
SELECT cnt
  FROM bar NATURAL JOIN foo
```

Lösung:

a) $\rho_{X(t1.a,t1.b,t1.c,t2.a,t2.b,t2.c)}(\rho_{t1}(\pi_{a,b,c}(R)) \times \rho_{t2}(\pi_{a,b,c}(R)))$

- `ORDER BY` geht nicht da Menge
- `WITH` geht nicht, deshalb zweimal bilden

b) $\pi_{L.a,L.b,R.a}(\sigma_{count(L.c)>5}(\Gamma(\sigma_{L.c>0}(L \bowtie_{L.a=R.b} R), (L.a, L.b, R.a), count, L.c)))$

- `IS NOT NULL` Check ist überflüssig, da `L.c>0` dies schon einschliesst

c) $\pi_{count(c)}((\Gamma(\sigma_{\text{c IS NOT NULL}}(S), (a,b), count, c)) \bowtie ((\sigma_{a>b}(\pi_{a,b}(R))) \cup (\sigma_{b>a}(\pi_{a,b}(R))))$

- UNION ALL erzeugt eine Multimengenrelation, die relationale Algebra kennt aber nur Mengen
- Temporärtabellen gibt es nicht in der relationalen Algebra Sieha auch 1a)
- count(*) als solches gibt es nicht. Weil man aber weiss, dass c nicht Null funktioniert Trick (Achtung: Γ und nicht nur γ)

Aufgabe 2
Berechnen Sie bei den folgenden relationalen Ausdrücken jeweils die Grössen (Anzahl der Zeilen und Spalten) **aller Resultat- und Zwischentabellen** in Abhängigkeit der Grössen der Ausgangstabellen. Falls keine genaue Grösse angebbar ist, begründen Sie kurz warum dies so ist und geben Sie eine möglichst genaue obere Abschätzung ab. Nutzen Sie für die Grössen folgende Angaben:

- n = Anzahl Tupel in L
- m = Anzahl Tupel in R
- s = Anzahl Attribute im Schema von L,
- t = Anzahl Attribute im Schema von R

a) $L \bowtie R$
b) $L \bowtie L$
c) $L \div R$
d) $\pi_{L.A,L.B}(L \times R)$
e) $\sigma_{L.A>R.A}\left(\rho_{L(A)}(R) \times R\right)$

Lösung:

a) Max. Anzahl Tupel: $n \cdot m$
Anzahl Attribute: $s + t - \#(gemeinsameAttribute)$

Konkrete Zahlen hängen ab von gemeinsamen Attributen und allenfalls gleichen Werten in denselbigen.

b) Max. Anzahl Tupel: n
Anzahl Attribute: s

Die Tupel mit NULL-Werten fallen raus

c) Max. Anzahl Tupel: $\frac{n}{m}$
Anzahl Attribute: $s - t$

Folgt alles direkt aus der Definition von $\div$

d) Max. Anzahl Tupel in $L \times R$: $n \cdot m$
Anzahl Attribute in $L \times R$: $s + t$

Max. Anzahl Tupel in $\pi_{L.A,L.B}(L \times R)$: n
Anzahl Attribute in $\pi_{L.A,L.B}(L \times R)$: 2

Da in der Projektion nur Attribute aus L vorkommen, fallen jeweils deren $m-1$ "Duplikate" weg (Mengen!)

e) Max. Anzahl Tupel in $\rho_{L(A)}(R)$: m
Anzahl Attribute in $\rho_{L(A)}(R)$: 1

Max. Anzahl Tupel in $\rho_{L(A)}(R) \times R$: m^2
Anzahl Attribute in $\rho_{L(A)}(R) \times R$: 2

Max. Anzahl Tupel in $\sigma_{L.A>R.A}\left(\rho_{L(A)}(R) \times R\right)$: $\frac{m \cdot (m-1)}{2}$
Anzahl Attribute in $\sigma_{L.A>R.A}\left(\rho_{L(A)}(R) \times R\right)$: 2

Anzahl Tupel: von den m Tupeln in L sind $m - 1$ Tupel kleiner als das grösste Tupel in R, $m - 2$ sind kleiner als das zweitgrösste etc.
$0 + 1 + 2 + \cdots + m - 1 = \frac{m \cdot (m-1)}{2}$