Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

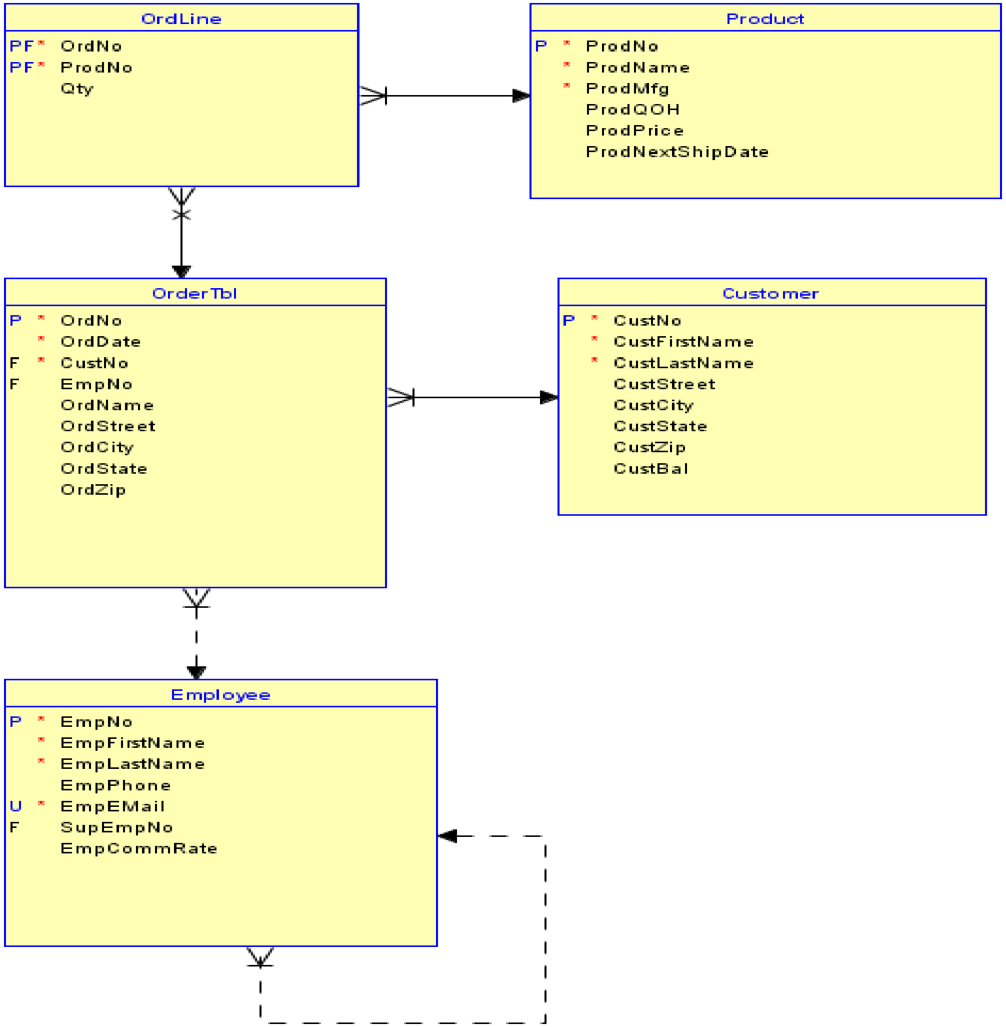
Module 6
SQL for Data Mining Input

Lesson 4: SQL Coding to Evaluate Association Rules

Lesson Objectives

- Define simple evaluation measures for association rules
- Write SELECT statements to generate and evaluate association rules with 2 items
- Use Common Table Expressions (CTEs)
- Discuss limitations of generating and evaluating association rules in SQL

Order Entry Tables (Operational Database)



OrdLine

PF* OrdNo
PF* ProdNo
Qty

Product

P * ProdNo
* ProdName
* ProdMfg
ProdQOH
ProdPrice
ProdNextShipDate

OrderTbl

P * OrdNo
* OrdDate
F * CustNo
F EmpNo
OrdName
OrdStreet
OrdCity
OrdState
OrdZip

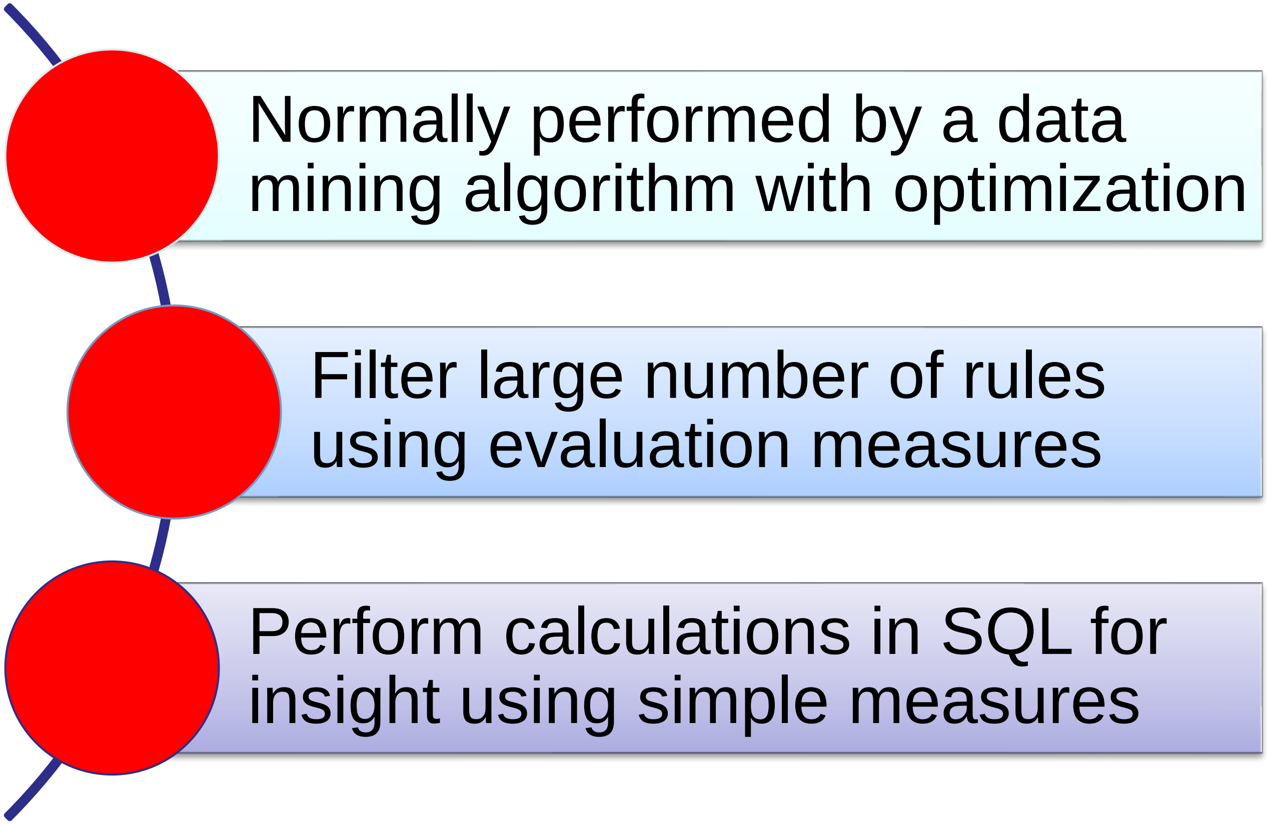
Customer

P * CustNo
* CustFirstName
* CustLastName
CustStreet
CustCity
CustState
CustZip
CustBal

Employee

P * EmpNo
* EmpFirstName
* EmpLastName
EmpPhone
U * EmpEMail
F SupEmpNo
EmpCommRate

Evaluation of Association Rules



Normally performed by a data mining algorithm with optimization

Filter large number of rules using evaluation measures

Perform calculations in SQL for insight using simple measures

Association Rule Evaluation

BasketId	*Items*
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

- All items means LHS and RHS
- Support: baskets with all items divided by total baskets
- Confidence: baskets with all items divided by baskets with LHS items alone
- Lift (importance measure): support of all items divided by LHS support times RHS support

Example

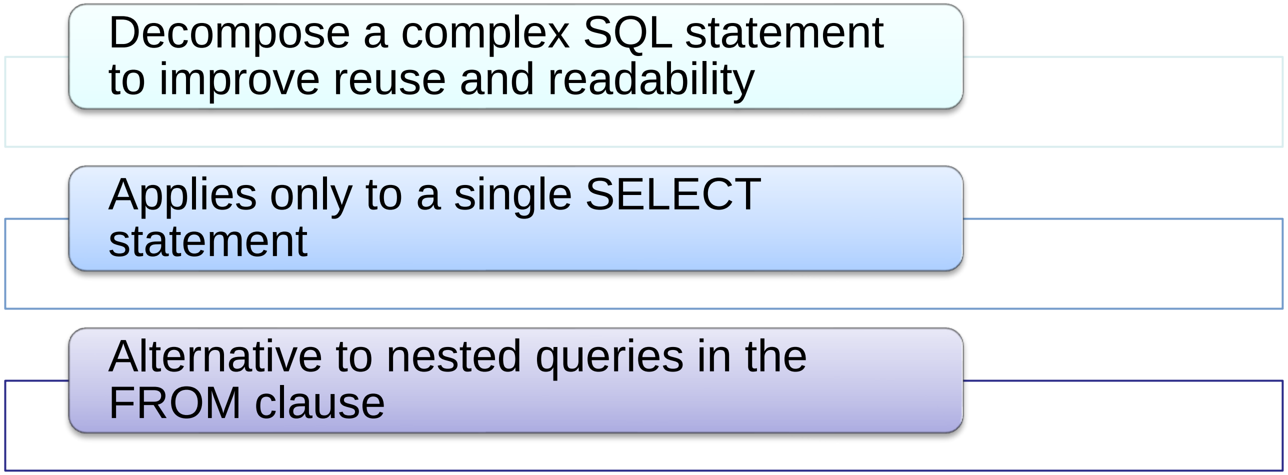
Diaper, Milk → Beer

Support: 0.4 (2/5)

Confidence: 0.66 (2/3)

Lift: 1.11 ((2/5) / ((3/5) * (3/5))

Common Table Expression (CTE)



Decompose a complex SQL statement to improve reuse and readability

Applies only to a single SELECT statement

Alternative to nested queries in the FROM clause

```
WITH CTEName1 AS
( <SELECTStatement> )
[, CTEName2 AS
( <SELECTStatment> ) … ,
   CTENamen AS
( <SELECTStatment> ) ]
<SELECTStatement> ;
```

SQL for Rule Evaluation I

```
-- Example 1 with Order Entry tables
-- 3 CTEs following WITH keyword separated by commas
WITH PairsCTE AS (
SELECT OL1.OrdNo, OL1.ProdNo LHSProd, OL2.ProdNo RHSProd
 FROM OrdLine OL1, OrdLine OL2
 WHERE OL1.OrdNo = OL2.OrdNo
   AND OL1.ProdNo <> OL2.ProdNo ),
RulesCTE AS (
SELECT LHSProd || ' -> ' || RHSProd as TheRule,
       LHSProd, RHSProd, COUNT(*) as SupportCnt
 FROM PairsCTE
 GROUP BY LHSProd, RHSProd ),
CountProductCTE AS (
SELECT ProdNo, COUNT(OrdNo) as ProductCount
 FROM OrdLine
 GROUP BY ProdNo )
```

SQL for Rule Evaluation II

```
-- Example 1 continued
-- SELECT statement using the CTEs
SELECT R.TheRule, R.SupportCnt,
   100.00 * (1.0 * R.SupportCnt / A.NumOrders )  AS SupportPercentage,
   100.00 * (1.0 * R.SupportCnt / C1.ProductCount ) AS Confidence,
   (1.0 * R.SupportCnt / A.NumOrders) /
  ((1.0 * C1.ProductCount / A.NumOrders) *
   (1.0 * C2.ProductCount / A.NumOrders))  AS Lift
 FROM RulesCTE R INNER JOIN CountProductCTE C1
   ON R.LHSProd = C1.ProdNo
   INNER JOIN CountProductCTE C2 ON R.RHSProd = C2.ProdNo
   CROSS JOIN
  ( SELECT COUNT(*) NumOrders FROM OrderTbl ) A
 ORDER BY Lift DESC, Confidence DESC;
```

Summary

- Definitions of basic rule evaluation measures
- SQL statement to generate association rules and basic evaluation measures
- CTE usage to simplify query formulation
- Usage of association rule mining algorithm for large scale data mining