2022.11.24

273 Session 2

Structured Query Language

Prof. Vibhanshu (Vibs) Abhishek

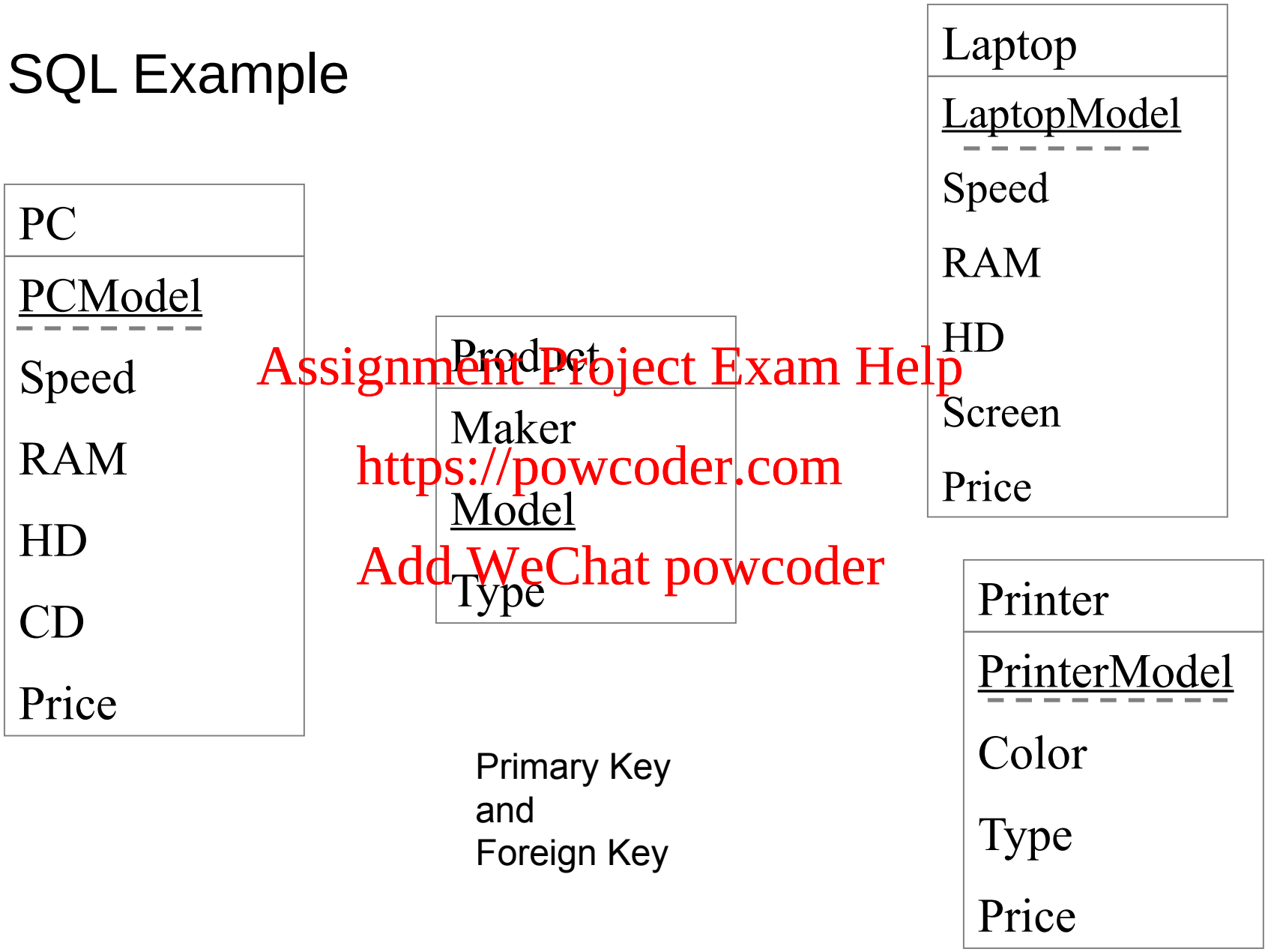
The Paul Merage School of Business

University of California, Irvine

Agenda

- Structured Query Language
  - Multi-table queries
- Reminders
  - Buy iClicker and register at iClicker.com

SQL Example



Joins - Queries From Multiple Tables

- Some queries require combining information from two or more tables.
- Example: I want the Model #, Manufacturer and Speed of all PCs. 'maker' is in the 'product' table, but 'speed' is in the 'pc' table.
- Example:


```
SELECT product.model,
       product.maker, pc.speed
FROM   product, pc
WHERE  product.model=pc.pcmodel;
```

Cartesian Product

- This helps to understand what a query from two tables does.
- Set theory concept - the Cartesian product of 2 sets: R and S is a set of pairs formed by choosing the an element from R and another from S.
- Example: R={a,b,c}, S={b,e}
  - R X S ={ a.b, a.e, b.b, b.e, c.b, c.e }

Cartesian Product Example

UCI_Students

SSN	L Name	F Name
000-00-0001	Doe	John
000-00-0002	Doe	Jason
000-00-0003	Smith	Jack

Football_Players

SSN	Team
000-00-0001	Chargers
111-11-1112	AntEaters
000-00-0003	49ers

- SELECT * FROM UCI_Students, Football_Players;
- SELECT UCI_Students.SSN, Football_players.SSN
  FROM UCI_Students, Football_Players;
- How can we get names of UCI students who also play football?

PRODUCT(maker, **model**, type)
PC(**PCmodel**, speed,ram, hd, cd, price)
LAPTOP(**Laptopmodel**,speed,ram,hd,screen,price)
PRINTER(**Printermodel**,color,type,price)

- List the model #, price of laptops made by maker 'Dell'.

- List the model, type and price of color printers from maker 'HP'

- List model, speed and price for PCs and Laptops that have the same price. (Do not JOIN on *model*)

Online help for learning SQL

- Online Practice :
  - http://www.w3schools.com/sql/trysql.asp?filename=trysql_select_all
- Orders(OrderID, CustomerID, EmployeeID, OrderDate, ShipperID)
- Employees(EmployeeID, LastName, FirstName, BirthDate, Photo, Notes)
- Customers(CustomerID, CustomerName, ContactName, Address, City, PostalCode, Country)
- List orderID, CustomerID and Employee first and last name where an employee placed an order for a customer.
- For above query also list customer name.

Sample Databases in MS Access

- For those of you who have Microsoft Access:
  - Posted 2 databases to Canvas:
  - Access database containing Computer store database example covered in class:
    product2.mdb (Links to an external site.)
  - Access database built by a student, containing (Book, Publisher) example and (Faculty, Course, Registration Student) example covered in class:
    Book Publisher and Faculty Course Registration Student DB.accdb

Practice

Book(ISBN, Title, PublisherID, Author, Cost, price)

Publisher(PublisherID, Name, Location)

1. List all books by the publisher with name 'Wiley and Sons'
2. List all book titles and publisher names
3. List the ISBN and the name of publisher where the title of the book is the same as the location of the publisher.

Grouping

- SELECT attribute list 1 FROM tables
  WHERE condition
  GROUP BY attribute list 2

- SELECT avg(price)
  From Product, PC
  Where ram>256 and product.model=pc.pcmodel
  Group By maker;

Group By .. Having

- SELECT attribute list1 FROM tables
  WHERE condition1
  GROUP BY attribute list2
  HAVING condition2;

- SELECT type, avg(price) FROM printer
  GROUP BY type
  HAVING min(price)>200;
- Condition2 applies to the group while condition1 applies to individual records.

Practice 2: Posted to Canvas

Student(StudentID, LastName, FirstName, Phone, Street, City, Zip)

Registration(**Student_ID, Course_Number, Term**, Reg_date, Score)

Course(Number, Term, Name, Description, Room_Number, Bldg_Name, **FacultyID**)

Faculty(FacultyID, Last_Name, First_Name, phone, Office_Room, Office_Bldg)

Foreign Keys in **Bold**

SQL Practice Set 2

- What was the average score for each course?
- What was the average score for each course with at least 50 students excluding students with a score greater than 100?
- List the names of all students that are being taught by Prof. Jack Smith.

Print the average price of PCs of each manufacturer that have more than 512 GB of RAM

A: SELECT AVG(price), FROM product, pc WHERE pc.ram>512;

B: SELECT AVG(price), FROM product, pc WHERE pc.ram>512 GROUP BY product.maker;

C: SELECT product.maker, AVG(price) FROM product, pc WHERE product.model=pc.pcmodel AND pc.ram>512 GROUP BY product.maker;

D: None of the above

E: All of the above

SQL Practice Set 3

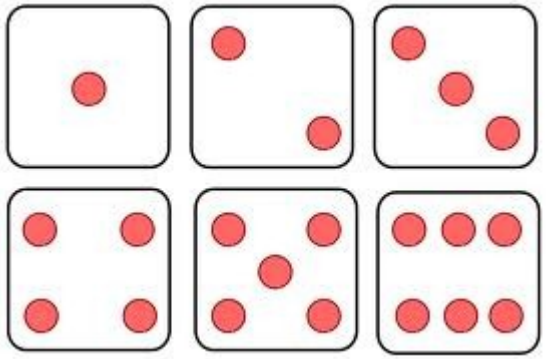
- Frequent_flier(Name, ffID, City, Street, zip, miles)
- Reservation(ResID, ffID, flightID, date, class, price, seatID)
- Flight(FlightID, Origin, Dest, DepTime, ArrTime, equipID)
- Equipment(EquipID, Desc, numFClass, numEClass)

Online help for learning SQL

- Online Practice :
  - http://www.w3schools.com/sql/trysql.asp?filename=trysql_select_all
- In OrderDetails table, for each productID, what is the average quantity ordered?
- OrderDetails(OrderDetailID, OrderID, ProductID, Quantity)

Probability Basics

Probability of an Event



- Denotes the likelihood of an event
  - Probability = 1 if the event is sure to happen
  - Probability = 0 if the event is sure not to happen
- Probability of obtaining a "head" when a fair coin is tossed
  - P(H) = ½
- Probability of getting an even number when rolling a die
  - P(E) = ½
- Probability of getting a number less than 3 (1 or 2) when rolling a die
  - P(L) = 1/3

Interpretation of Probability

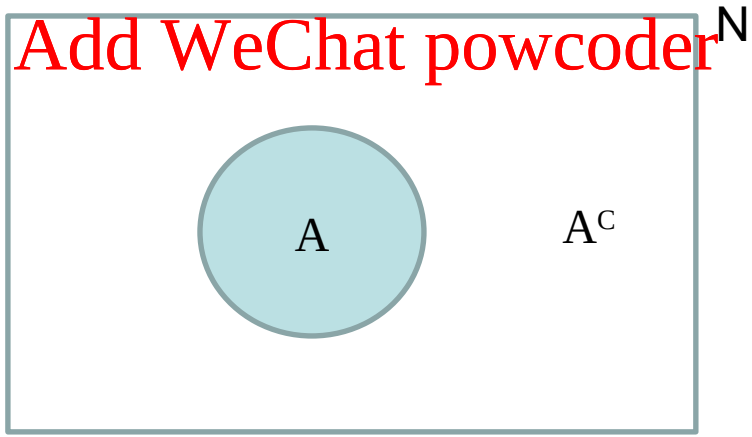
- **Frequentist**
  - The relative frequency is the probability
- You toss a coin 1000 times; *N=1000*
- You get 509 heads; *N(H) = 509*
- Pr(H) = *N(H)/N = 509/1000 = 0.509*

Combining Events

- Complement of the event A
  - Happens whenever A does not happen
- Union of events A and B
  - Happens whenever either A or B or both events happen
- Intersection of A and B
  - Happens whenever both A and B happen
- Conditional Probability of A Given B
  - The updated probability of A, possibly changed to reflect the fact that B has happened

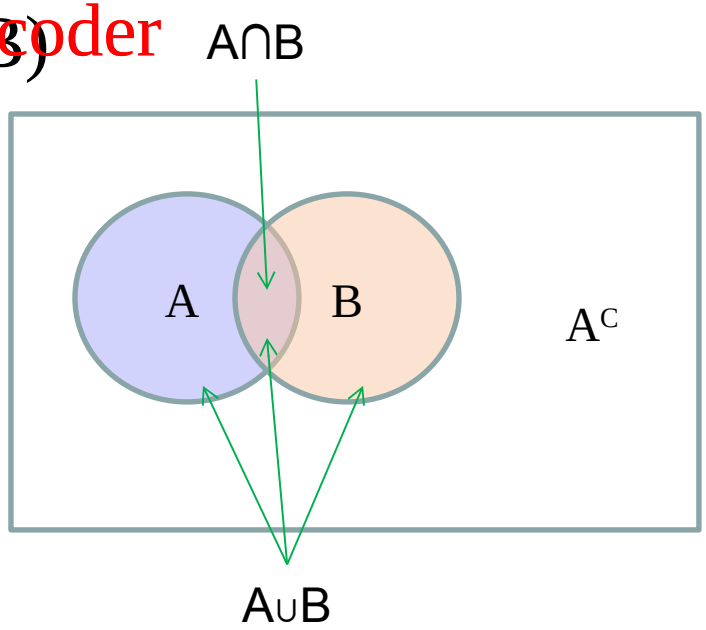
Complement of an Event

- The event “not A” (A^C— read A complement) happens whenever A does not
- $P(A^C) = 1 - P(A)$



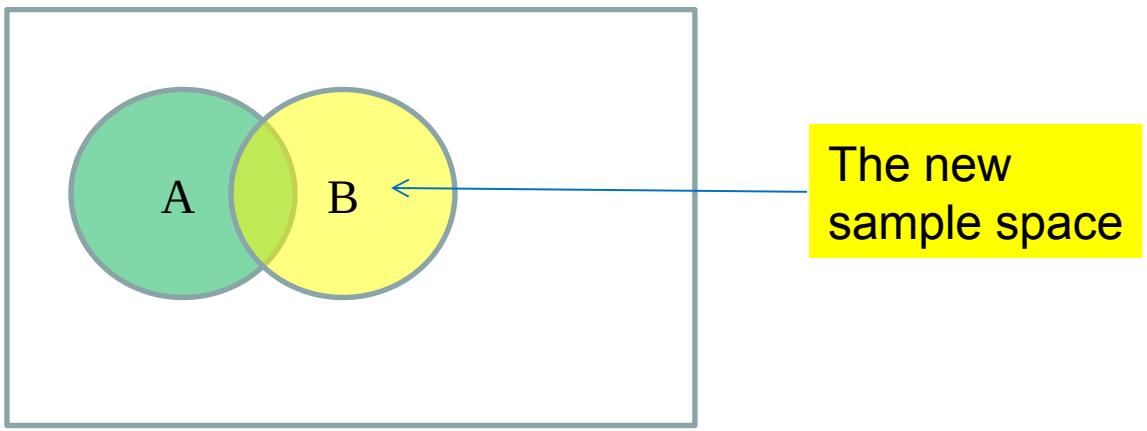
Union and Intersection of Events

- Union happens whenever at least one happens
  - $A \cup B$ happens if A happens, B happens, or both happen
- Intersection happens whenever both happen
  - $A \cap B$ happens if A and B both happen
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cap B) = P(A) + P(B) - P(A \cup B)$



Conditional Probability

- P(A|B)
  - Probability of event A given that event B has happened
  - In short, probability of A given B
- If B has already occurred the sample space reduces to only the outcomes associated with B
- P(A|B) = N(A∩B)/N(B) = P(A∩B)/P(B)

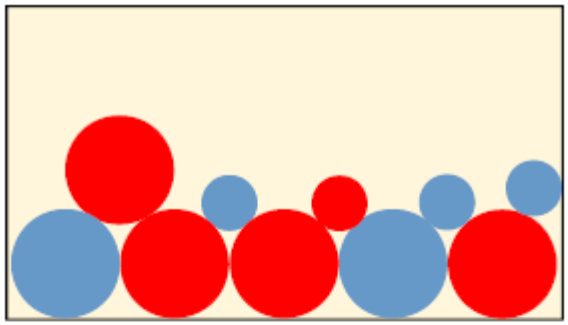


Independent Event

- Two events A and B are independent if one event has no influence on the other
  - $P(A|B) = P(A)$
  - This is also equivalent to:
    - $P(A \cap B) = P(A) \times P(B)$
    - $P(B|A) = P(B)$
- If two events are not independent, then they are dependent

Size and color are not independent

Example



Consider a box with 10 balls

If a ball is picked randomly

- Pr[Large] = 6/10 = 0.6
- Pr[Blue] = 5/10 = 0.5
- Pr[Large & Red] = 0.4
- Pr[Large OR Red]

  = Pr[Large] + Pr[Red] – Pr[Large & Red]

  = 0.6 + 0.5 – 0.4 = 0.7
- Pr[Large OR Red]

  = 1 – Pr[Small & Blue] = 1 – 0.3 = 0.7
- Pr[Large|Red]

  = N[Large & Red]/N[Red] = 4/5 = 0.8

	Red	Blue	
Large	4	2	*6*
Small	1	3	*4*
	5	*5*	***10***

	Red	Blue	
Large	0.4	0.2	*0.6*
Small	0.1	0.3	*0.4*
	0.5	*0.5*	***1.0***

Bayes' Rule

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A) \times P(A)}{P(B)}$$

- P(Cancer|Smoker)= ___?
  - C=Cancer, S=Smoker
  - 20% of cancer patients are (have been) smokers
    - P(Smoker|Cancer) = P(S|C) = 0.2
  - Only 2% of the overall population have cancer
    - P(Cancer) = P(C) = 0.02
  - 10% of the overall population smoke
    - P(Smoker) = P(S) = 0.1
- – P(Cancer|Smoker] = 0.2×0.02/0.1 = **0.04**

Loan Application Data

Income	CreditRating	Liability	Default	Approve
high	excellent	normal	true	yes
high	excellent	normal	false	yes
low	excellent	normal	true	no
medium	good	normal	true	no
medium	poor	high	true	no
medium	poor	high	false	yes
low	poor	high	false	no
high	good	normal	true	yes
high	poor	high	true	no
medium	good	high	true	no
high	good	high	false	no
low	good	normal	false	no
low	excellent	high	true	no
medium	good	nomal	false	yes

Loan Application Data

- N[Income=high] = ___
- N[CreditRating=poor] = ___
- N[Income=high, Liability=normal] = ___
- P[Income=high] = ___
- P[CreditRating=poor] = ___
- P[Income=high, Liability=normal] = ___
- P[Liability=normal | Income=high] = ___

Contingency Table
(Expressing relationship between two attributes)

		Liability		
		normal	high	**Total**
CreditRating	excellent	3	1	**4**
	good	4	2	**6**
	poor	0	4	**4**
	Total	**7**	**7**	**14**

- N[CreditRating=excellent] = ___
- N[Liability=normal] = ___
- N[CreditRating=good, Liability=normal] = ___

- P[CreditRating=excellent] = ___
- P[Liability=normal] = ___
- P[CreditRating=good, Liability=normal] = ___
- P[CreditRating=good | Liability=normal] = ___

P[CreditRating=excellent | Liability=high] =

A: 1/4

B: 1/7

C: 1/2

D: 1/6

E: 1/3

RFM, Pivot Tables and London Jets Data

- http://www.dbmarketing.com/articles/Art149.htm
- London Jets Data in Excel format posted on Canvas for RFM analysis and Pivot tables.
  - Do RFM analysis on this data
  - Think about strategies that London Jets could use to revive their fortunes
- Go to http://office.microsoft.com/en-us/
  - Search for “Pivot Table” and read up on creating and using them

Next Session

- Information Theory
- Working with datasets