BANA 273 Session 1

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Data Mi teligence

Prof. Vibhanshu (ishek

The Paul Merage School of Business

University of California, Irvine

Agenda

- Introduction
  - Instructor and TA
- Course Logi
- Data Mining
- SQL

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

About the Instructor

- Undergraduate degree in Computer Sc & Engr
  - Indian Institute of Technology, Kanpur
- Masters in Statistics
  - University rton School
- Ph.D in Ope n Management
  - University of Pennsylvania on School
- Assistant Professor at Carnegie Mellon University
- Just joined UCI ☺

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Teaching

- 273 Business Intelligence for Analytical Decision Making – BANA/FEMBA
- 173 Busines ytical Decision Making – U
- Digital Marketing Analyti
- Managing Disruptive Technologies

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Research

- Effect of emerging technologies on consumers, firms and markets
- Digital Market

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Some Consulting Projects

- Sequoia Capital
- Adobe
- McKinsey
- LEGO
- Kaplan
- One Championship
- Rivigo

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

BSA Survey on Analytics

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Other than work …

- Startups
- Photography
- Travel
- Anvi

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Course Topics

- Getting data from Database: Structured Query Language
- RFM Analysis, Excel Pivot Tables
- Basic Probability & Information Theory
- Preparing Data
- Data Mining Methods
  - Classification Using Naïve Bayes
  - Classification using Decision Trees
  - Association Rule Mining
  - Clustering

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Who should take this class

- Required core class for MSBA students
- No prior technical knowledge is assumed
- Aptitude for learning technical concepts
  - Algorithms
  - Software (MS Excel, Weka
  - Understand business conce
- Course is recommended for students interested in understanding the techniques and applications of data mining and acquiring hands-on skills for making intelligent business decisions in data-rich organizations.

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Guidelines for this course

- No cell phone or laptop usage in-class unless required for class
- UCI Office hours. Do not
  - Post class sites
  - Do not share individual ass
- Learn about US businesses, language and culture
- Groups for class project

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Motivation: Business Value of Customer Data

- What can a retail store do with shopping basket data?

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

How can big data be used to reduce fuel consumption?

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

© Prof. V Abhishek,

How data is utilized is key

- Accurate, timely intelligence
- Rapid deploy
- Effective tacti
- Relentless follow-up and as

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

© Prof. V Choudhary,

Business Intelligence & Data Mining

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Business Intelligence: Components

- Reporting
  - Tooth-paste sales in September?
  - Cosmetics

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Business Intelligence

- Reporting
  - Tooth-paste sales in September?
  - Cosmetics
- Dashboards
  - Easy access to performance
  - Ability to drill down into the data

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Example of Dashboard from MicroStrategy

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Example of Dashboard from MicroStrategy

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Business Intelligence

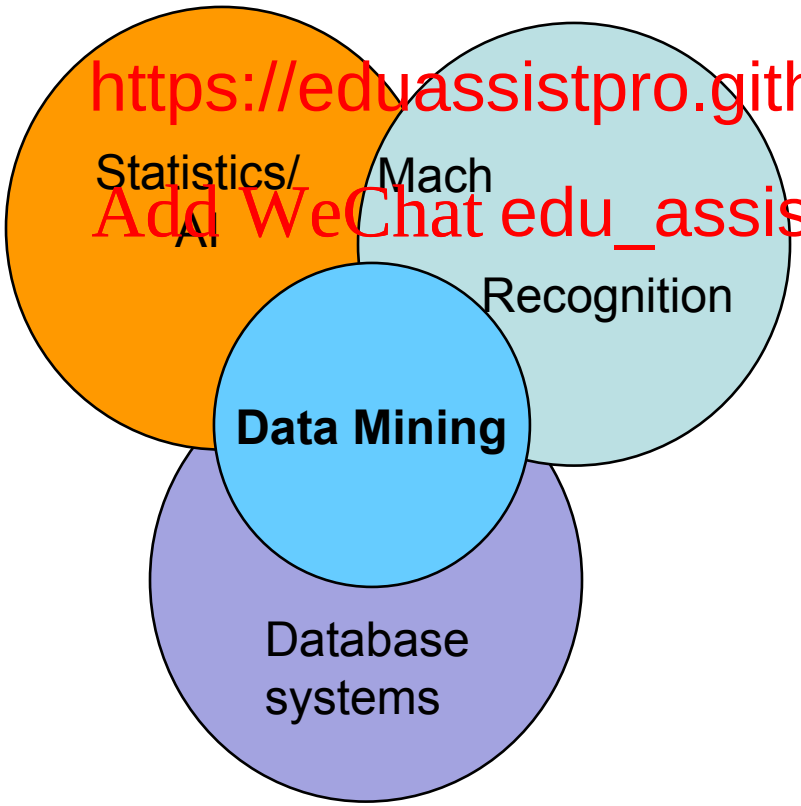
- Reporting
  - Tooth-paste sales in September?
  - Cosmetics
- Dashboards
  - Easy access to performance
  - Ability to drill down into the data
- Data Mining

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

What Is Data Mining?

*Data mining is a **process** of discovering and interpreting (unknown) patterns in data to solve a business problem.*

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro



Data Sources for Data Mining

- Public and private data
  - Demographic data
  - Customer data
  - Transactio
  - Product data
- Accounting and marketin
  - Web logs (click-stream data)
  - Application server-logs
  - Third party server-logs
- Data integration is an important challenge

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Who Uses Data Mining?

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Amazon's Personalized Web Pages

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Who Uses Data Mining?

- Firms with corporate offices in OC

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Data Mining Methods...

- Classification
- Clustering
- Association Rule Discovery
- Decision Tree

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

- Visualization
- Collaborative filtering
- Text Mining
- Neural networks and Deep learning

Examples of Visualizing Data: Word Cloud

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Classification: Definition

- Given a collection of records (*training set*)
  - Each record contains a set of *attributes*, one of the attributes is the *class*.
- Find a *mo* is a function of the values of other attrib
- Goal: previously unseen ould be assigned a class as accurately as possible.

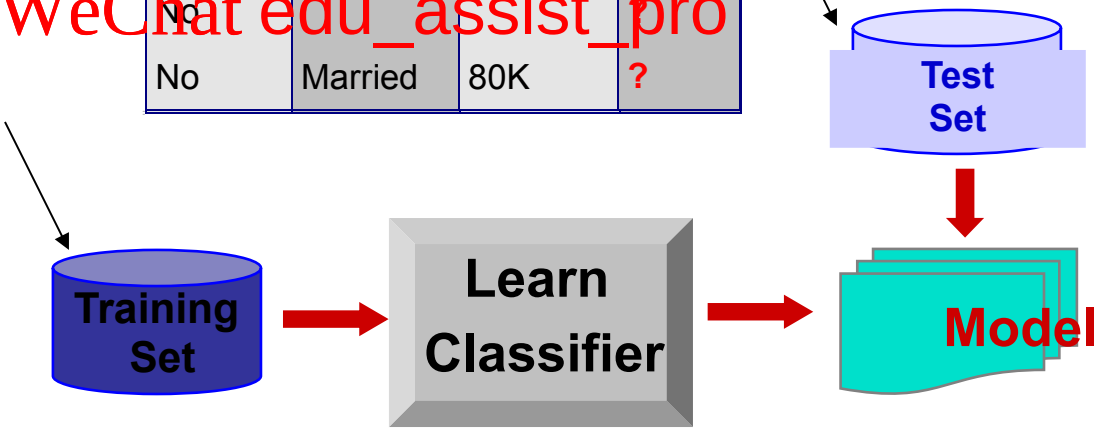
Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Classification Example

Tid	Refund (categorical)	Marital Status (categorical)	Taxable Income (continuous)	Cheat (class)
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

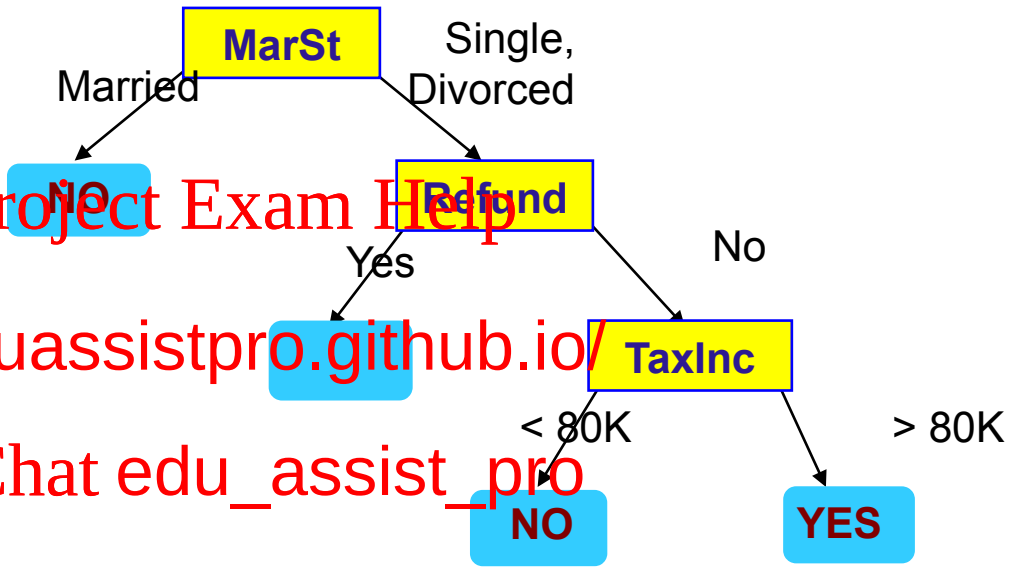
Refund	Marital Status	Taxable Income	Cheat
No	Single	75K	?
		K	?
		0K	?
Yes			?
No			?
No	Married	80K	?

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro



Example of Decision Tree

Tid	Refund	Marital Status	Taxable Income	Cheat
	categorical	categorical	continuous	class
1	Yes	Single	125K	No
2	No	Married	100K	
3	No	Single	70K	
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

(11, Yes, Divorced, 115K)

(12, No, Single, 150K)

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

© Prof. V Choudhary,

Other Applications for Classification

- Direct Marketing – LL Bean
  - Goal: Reduce cost of mailing by *targeting* a set of consumers likely to buy a new cell-phone product.

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Other Applications for Classification

- Customer Attrition/Churn at AT&T
  - 16% in 2013
  - Goal: To predict whether a customer is likely to be lost to a compe

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Who is Watson?

Assignment Project Exam Help

IBM Watson: https://eduassistpro.github.io/

https://www.youtube.com/watch?v=_I-M7O_bRNg

Add WeChat edu_assist_pro

0 to 0:25; 1:10 to 2:03; 3:10 to 4:27; 6:30 to 7:40

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

https://www.riverscasino.com/pittsburgh/BrainsVsAI

Assignment Project Exam Help
Str age
https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Database Management Systems *(DBMS)*

- Software through which users and application programs interact with a database

Assignment Project Exam Help

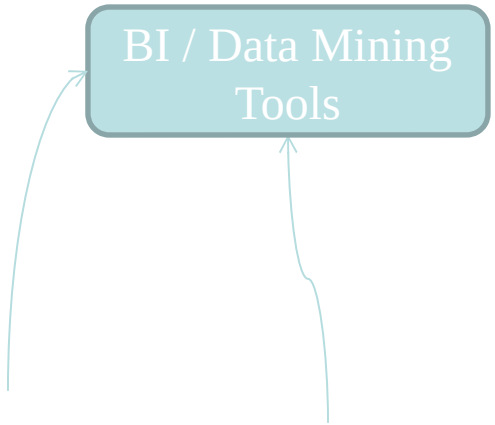
https://eduassistpro.github.io/

Add WeChat edu_assist_pro

DBMS DBMS

DBMS manages data resources like an operating system manages hardware resources

Data Warehouse

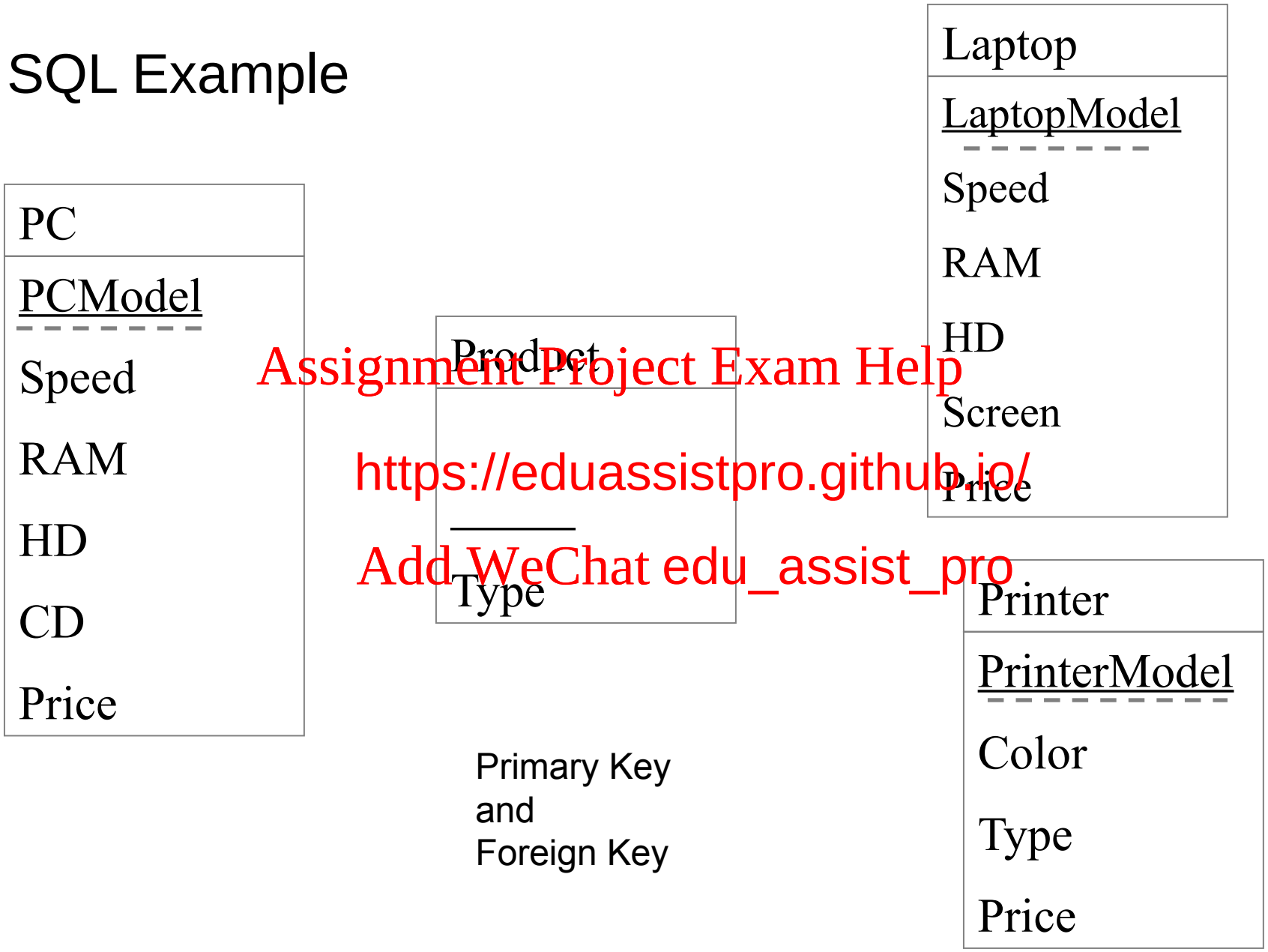


Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

SQL Example



Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Sample data for table PC					
PCmodel	**speed**	**ram**	**hd**	**cd**	**price**
1001	133	16	1.6	6x	1595
1002	120	16	1.6	6x	1399
1003	166	24	2.5	6x	1899
1004	166	32	2.5	8x	1999
1005	16			8x	1999
1006	20				2099
1007	200	32			2349
1008	180	32	2	8x	3699
1009	200	32	2.5	8x	2599
1010	160	16	1.2	8x	1495

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Sample data for PRODUCT		
maker	**model**	**type**
A	1001	pc
A	1002	pc
A	1003	pc
B	1004	pc
B	1006	
B	3002	printer
B	3004	printer
C	1005	pc
C	1007	pc
D	1008	pc
D	1009	pc
D	1010	pc
D	2001	laptop
D	2002	laptop
D	2003	laptop
	3001	printer
	3003	printer
	2004	laptop
	2008	laptop
F	2005	laptop
G	2006	laptop
G	2007	laptop
H	3005	printer
I	3006	printer

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Sample data for table LAPTOP					
LaptopModel	**speed**	**ram**	**hd**	**screen**	**price**
2001	100	20	1.1	9.5	1999
2002	117	12	0.75	11.3	2499
2003				10.4	3599
2004				11.2	3499
2005	133	16	1.1	11.3	2599
2006	120	8	0.81	12.1	1999
2007	150	16	1.35	12.1	4799
2008	120	16	1.1	12.1	2099

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Sample data for table PRINTER			
PrinterModel	**color**	**type**	**price**
3001	TRUE	ink jet	275
3002		jet	269
3003	FALSE	r	829
3004	FALSE	r	879
3005	FALSE	ink jet	180
3006	TRUE	dry	470

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

The SELECT Statement

- Used for queries on single or multiple tables
  Clauses of the SELECT statement:
  - SELECT - List the columns (and expressions) that should be returned from the query
  - FROM - Indicate the table(s) or view(s) from which data will be obtained
  - WHERE - Indica row will be included in the result
    Optional clauses:
    - GROUP BY - Indicate categorizatio
    - HAVING - Indicate the conditions under which a category (group) will be included
    - ORDER BY - Sorts the result according to specified criteria

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

SQL Query

Printer Table

printermodel	color	type	price
3004	FALSE	laser	879
3005	FALSE	ink jet	180
3006	TRUE	dry	470

Assignment Project Exam Help

- SELECT <attri t>;
- SELECT printe ter;
- SELECT <attribute list> FRO WHERE <condition>;
- SELECT printermodel, color FROM printer WHERE price<300;

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Aggregation Operators

- SUM
  - SELECT SUM(price) FROM PC;
- AVG
  - SELECT AVG(price) FROM PC WHERE speed>800;
- MIN
  - SELECT MIN(pric
- MAX
  - SELECT MAX(price) FROM PC WHE
- COUNT
  - SELECT COUNT(model) FROM PC;
- Keyword DISTINCT (not an aggregation operator)
  - SELECT DISTINCT type FROM printer;

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Practice

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

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

- List the model and speed of P ice<2000.
- How many laptops are there with screen size>11.
- What is average price of laptops with ram=20.
- List all attributes of all color printers

Practice

Book(ISBN, Title, Publisher ID, Author, Cost, price)

Publisher(Publisher ID, Name, Location)

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

1. List the publishers and their location.
2. List all book titles f
3. How many books e ook has one entry in the book table)
4. What is the total cost of the books?
5. List all distinct book titles
6. What is the expected profit per book for books published by Publisher ID 222?

Print the speed and price of PCs that cost less than $2000

A: Select speed, price from printer where price < 2000

B: Select speed, price from laptop where price < 2000

C: Select spee < 2000

D: Select spee here price < 2000

E: None of the above

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Print the price of the cheapest laptop

A: Select price from printer

B: Select min(price) from pc

C: Select min(

D: Select chea

E: None of the above

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Online help for learning SQL

- Tutorials:
- http://sqlzoo.net/wiki/SQL_Tutorial
- Practice with data:
- http://www.w3schools.com/sql/trysql.asp?filename=trysql_select_all

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

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
FROM   product, pc
WHERE  product.model=pc.pcmodel;
```

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Cartesian Product

- This helps to understand what a query from two tables does.
- Set theory con duct of 2 sets: R and S is a set o sing the an element from R and another
- Example: R={a,b,c}, S={b,e}
  - R X S ={ a.b, a.e, b.b, b.e, c.b, c.e }

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

Cartesian Product Example

UCI_Students

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

Football_Players

SSN	Team
000-00-0001	Chargers
111-11-1112	AntEaters
03	49ers

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

- SELECT * FROM UCI_Students, Footb
- 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)

Assignment Project Exam Help

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

https://eduassistpro.github.io/

- List the model, type and price of PC from maker 'G'

Add WeChat edu_assist_pro

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

© Prof. V Choudhary,

Practice

Book(ISBN, Title, Publisher ID, Author, Cost, price)

Publisher(Publisher ID, Name, Location)

1. List all books by the publisher with name 'Wiley and Sons'
2. List all book titles a
3. List the ISBN and t e of the book is the same as the location of the publisher.

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

© Prof. V Choudhary,

Practice 2

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

Registration(**Studen** **Term**, Reg_date, Score)

Course(Number, Te om_Number,
Bldg_Name, **Faculty_ID**)

Faculty(Faculty_ID, Last_Name, Fir one, Office_Room,
Office_Bldg)

Foreign Keys in **Bold**

Assignment Project Exam Help
https://eduassistpro.github.io/
Add WeChat edu_assist_pro

© Prof. V Choudhary,

Next Session

- SQL Continued

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

© Prof. V Choudhary,