SQL for Data Analysis

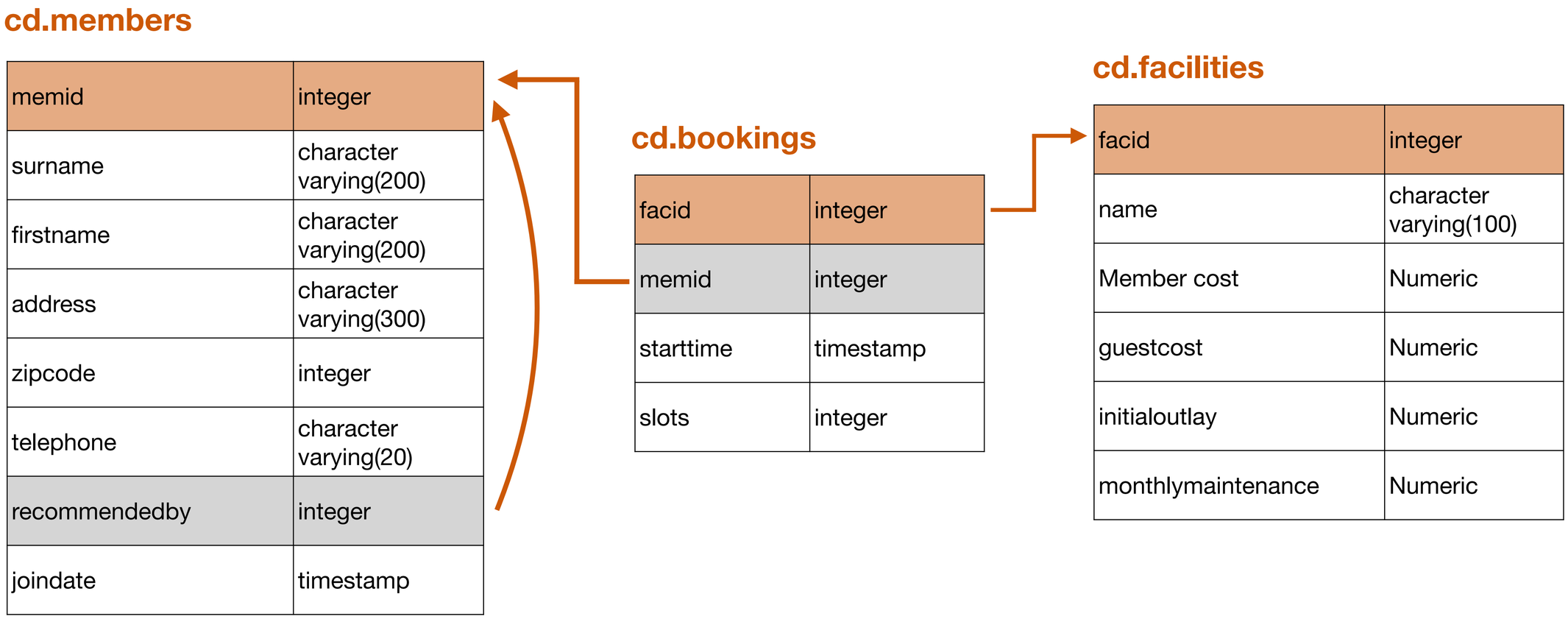
PRACTICE EXERCISES ON CLUB DATA

BASIC QUERIES

SQL Practice Exercises on Club Data

BI DATA LABS

Club Data Db Schema



BI DATA LABS

Retrieve specific columns from a table

QUESTION

How can you retrieve all the information from the cd.facilities table?

EXPECTED RESULT

facid	name	membercost	guestcost	initialoutlay	monthlymaintenance
0	Tennis Court 1	5	25	10000	200
1	Tennis Court 2	5	25	8000	200
2	Badminton Court	0	15.5	4000	50
3	Table Tennis	0	5	320	10
4	Massage Room 1	35	80	4000	3000
5	Massage Room 2	35	80	4000	3000
6	Squash Court	3.5	17.5	5000	80
7	Snooker Table	0	5	450	15
8	Pool Table	0	5	400	15

Retrieve specific columns from a table

QUESTION

You want to print out a list of all of the facilities and their cost to members. How would you retrieve a list of only facility names and costs?

EXPECTED RESULT

name	membercost
Tennis Court 1	5
Tennis Court 2	5
Badminton Court	0
Table Tennis	0
Massage Room 1	35
Massage Room 2	35
Squash Court	3.5
Snooker Table	0
Pool Table	0

Control which rows are retrieved

QUESTION

How can you produce a list of facilities that charge a fee to members?

EXPECTED RESULT

facid	name	membercost	guestcost	initialoutlay	monthlymaintenance
0	Tennis Court 1	5	25	10000	200
1	Tennis Court 2	5	25	8000	200
4	Massage Room 1	35	80	4000	3000
5	Massage Room 2	35	80	4000	3000
6	Squash Court	3.5	17.5	5000	80

Control which rows are retrieved - part 2

QUESTION

How can you produce a list of facilities that charge a fee to members, and that fee is less than 1/50th of the monthly maintenance cost?

Return the facid, facility name, member cost, and monthly maintenance of the facilities in question.

EXPECTED RESULT

facid	name	membercost	monthlymaintenance
4	Massage Room 1	35	3000
5	Massage Room 2	35	3000

Control which rows are retrieved - part 2

QUESTION

How can you produce a list of facilities that charge a fee to members, and that fee is less than 1/50th of the monthly maintenance cost?

Return the facid, facility name, member cost, and monthly maintenance of the facilities in question.

EXPECTED RESULT

facid	name	membercost	monthlymaintenance
4	Massage Room 1	35	3000
5	Massage Room 2	35	3000

Basic string searches

QUESTION

How can you produce a list of all facilities with the word 'Tennis' in their name?

EXPECTED RESULT

facid	name	membercost	guestcost	initialoutlay	monthlymaintenance
0	Tennis Court 1	5	25	10000	200
1	Tennis Court 2	5	25	8000	200
3	Table Tennis	0	5	320	10

Matching against multiple possible values

QUESTION

How can you retrieve the details of facilities with ID 1 and 5? Try to do it without using the OR operator.

EXPECTED RESULT

facid	name	membercost	guestcost	initialoutlay	monthlymaintenance
1	Tennis Court 2	5	25	8000	200
5	Massage Room 2	35	80	4000	3000

Classify results into buckets

QUESTION

How can you produce a list of facilities, with each labelled as 'cheap' or 'expensive' depending on if their monthly maintenance cost is more than $100?

Return the name and monthly maintenance of the facilities in question.

EXPECTED RESULT

name	cost
Tennis Court 1	expensive
Tennis Court 2	expensive
Badminton Court	cheap
Table Tennis	cheap
Massage Room 1	expensive
Massage Room 2	expensive
Squash Court	cheap
Snooker Table	cheap
Pool Table	cheap

Working with Dates

QUESTION

How can you produce a list of members who joined after the start of September 2012?

Return the memid, surname, firstname, and joindate of the members in question.

EXPECTED RESULT

memid	surname	firstname	joindate
24	Sarwin	Ramnaresh	2012-09-01 08:44:42
26	Jones	Douglas	2012-09-02 18:43:05
27	Rumney	Henrietta	2012-09-05 08:42:35
28	Farrell	David	2012-09-15 08:22:05
29	Worthington-Smyth	Henry	2012-09-17 12:27:15
30	Purview	Millicent	2012-09-18 19:04:01
33	Tupperware	Hyacinth	2012-09-18 19:32:05
35	Hunt	John	2012-09-19 11:32:45
36	Crumpet	Erica	2012-09-22 08:36:38
37	Smith	Darren	2012-09-26 18:08:45

Removing duplicates, and ordering results

QUESTION

How can you produce an ordered list of the first 10 surnames in the members table?

The list must not contain duplicates.

EXPECTED RESULT

surname
Bader
Baker
Boothe
Butters
Coplin
Crumpet
Dare
Farrell
GUEST
Genting

Combining results from multiple queries

QUESTION

You, for some reason, want a combined list of all surnames and all facility names. Yes, this is a bit of a contrived example :-).

Produce that list!

EXPECTED RESULT

surname
Tennis Court 2
Worthington-Smyth
Badminton Court
Pinker
Dare
Bader
Mackenzie
Crumpet
Massage Room 1
Squash Court

Simple aggregation

QUESTION

You'd like to get the signup date of your last member. How can you retrieve this information?

EXPECTED RESULT

latest
2012-09-26 18:08:45

More aggregation

QUESTION

You'd like to get the first and last name of the last member(s) who signed up - not just the date.

How can you do that?

EXPECTED RESULT

firstname	surname	joindate
Darren	Smith	2012-09-26 18:08:45

AGGREGATION

SQL Practice Exercises on Club Data

BI DATA LABS

Count the number of facilities

QUESTION

For our first foray into aggregates, we're going to stick to something simple.

We want to know how many facilities exist - simply produce a total count.

EXPECTED RESULT

count
9

Count the number of expensive facilities

QUESTION

Produce a count of the number of facilities that have a cost to guests of 10 or more.

EXPECTED RESULT

count
6

Count the number of recommendations each member makes

QUESTION

Produce a count of the number of recommendations each member has made. Order by member ID.

EXPECTED RESULT

recommendedby	count
1	5
2	3
3	1
4	2
5	1
6	1
9	2
11	1
13	2
15	1

List the total slots booked per facility

QUESTION

Produce a list of the total number of slots booked per facility. For now, just produce an output table consisting of facility id and slots, sorted by facility id.

EXPECTED RESULT

facid	Total Slots
0	1320
1	1278
2	1209
3	830
4	1404
5	228
6	1104
7	908
8	911

List the total slots booked per facility in a given month

QUESTION

Produce a list of the total number of slots booked per facility in the month of September 2012.

Produce an output table consisting of facility id and slots, sorted by the number of slots.

EXPECTED RESULT

facid	Total Slots
5	122
3	422
7	426
8	471
6	540
2	570
1	588
0	591
4	648

List the total slots booked per facility per month

QUESTION

Produce a list of the total number of slots booked per facility per month in the year of 2012.

Produce an output table consisting of facility id and slots, sorted by the id and month.

EXPECTED RESULT

facid	month	Total Slots
0	7	270
0	8	459
0	9	591
1	7	207
1	8	483
1	9	588
2	7	180
2	8	459
2	9	570
3	7	104

Find the count of members who have made at least one booking

QUESTION

Find the total number of members (including guests) who have made at least one booking.

EXPECTED RESULT

count
30

List facilities with more than 1000 slots booked

QUESTION

Produce a list of facilities with more than 1000 slots booked.

Produce an output table consisting of facility id and slots, sorted by facility id.

EXPECTED RESULT

facid	Total Slots
0	1320
1	1278
2	1209
4	1404
6	1104

Find the total revenue of each facility

QUESTION

Produce a list of facilities along with their total revenue.

The output table should consist of facility name and revenue, sorted by revenue.

Remember that there's a different cost for guests and members!

EXPECTED RESULT

name	revenue
Table Tennis	180
Snooker Table	240
Pool Table	270
Badminton Court	1906.5
Squash Court	13468.0
Tennis Court 1	13860
Tennis Court 2	14310
Massage Room 2	15810
Massage Room 1	72540

Find facilities with a total revenue less than 1000

QUESTION

Produce a list a of facilities with a total revenue less than 1000.

Produce an output table consisting of facility name and revenue, sorted by revenue.

Remember that there's a different cost for guests and members!

EXPECTED RESULT

name	revenue
Table Tennis	180
Snooker Table	240
Pool Table	270

Output the facility id that has the highest number of slots booked

QUESTION

Output the facility id that has the highest number of slots booked. For bonus points, try a version without a LIMIT clause.

This version will probably look messy!

EXPECTED RESULT

facid	Total Slots
4	1404

BI DATA LABS

List the total slots booked per facility per month, part 2

QUESTION

Produce a list of the total number of slots booked per facility per month in the year of 2012.

In this version, include output rows containing totals for all months per facility, and a total for all months for all facilities.

The output table should consist of facility id, month and slots, sorted by the id and month. When calculating the aggregated values for all months and all facids, return null values in the month and facid columns.

EXPECTED RESULT

facid	month	slots
0	7	270
0	8	459
0	9	591
0		1320
1	7	207
1	8	483
1	9	588
1		1278
2	7	180

List the total hours booked per named facility

QUESTION

Produce a list of the total number of hours booked per facility, remembering that a slot lasts half an hour.

The output table should consist of the facility id, name, and hours booked, sorted by facility id.

Try formatting the hours to two decimal places.

EXPECTED RESULT

facid	name	Total Hours
0	Tennis Court 1	660.00
1	Tennis Court 2	639.00
2	Badminton Court	604.50
3	Table Tennis	415.00
4	Massage Room 1	702.00
5	Massage Room 2	114.00
6	Squash Court	552.00
7	Snooker Table	454.00
8	Pool Table	455.50

List each member's first booking after September 1st 2012

QUESTION

Produce a list of each member name, id, and their first booking after September 1st 2012. Order by member ID.

EXPECTED RESULT

surname	firstname	memid	starttime
GUEST	GUEST	0	2012-09-01 08:00:00
Smith	Darren	1	2012-09-01 09:00:00
Smith	Tracy	2	2012-09-01 11:30:00
Rownam	Tim	3	2012-09-01 16:00:00
Joplette	Janice	4	2012-09-01 15:00:00
Butters	Gerald	5	2012-09-02 12:30:00
Tracy	Burton	6	2012-09-01 15:00:00
Dare	Nancy	7	2012-09-01 12:30:00
Boothe	Tim	8	2012-09-01 08:30:00

Produce a list of member names, with each row containing the total member count

QUESTION

Produce a list of member names, with each row containing the total member count.

Order by join date, and include guest members.

EXPECTED RESULT

count	firstname	surname
31	GUEST	GUEST
31	Darren	Smith
31	Tracy	Smith
31	Tim	Rownam
31	Janice	Joplette
31	Gerald	Butters
31	Burton	Tracy
31	Nancy	Dare
31	Tim	Boothe
31	Ponder	Stibbons
31	Charles	Owen

Produce a numbered list of members

QUESTION

Produce a monotonically increasing numbered list of members (including guests), ordered by their date of joining.

Remember that member IDs are not guaranteed to be sequential.

EXPECTED RESULT

row_number	firstname	surname
1	GUEST	GUEST
2	Darren	Smith
3	Tracy	Smith
4	Tim	Rownam
5	Janice	Joplette
6	Gerald	Butters
7	Burton	Tracy
8	Nancy	Dare
9	Tim	Boothe
10	Ponder	Stibbons

Output the facility id that has the highest number of slots booked, again

QUESTION

Output the facility id that has the highest number of slots booked. Ensure that in the event of a tie, all tieing results get output.

EXPECTED RESULT

facid	total
4	1404

Rank members by (rounded) hours used

QUESTION

Produce a list of members (including guests), along with the number of hours they've booked in facilities, rounded to the nearest ten hours.

Rank them by this rounded figure, producing output of first name, surname, rounded hours, rank. Sort by rank, surname, and first name.

EXPECTED RESULT

firstname	surname	hours	rank
GUEST	GUEST	1200	1
Darren	Smith	340	2
Tim	Rownam	330	3
Tim	Boothe	220	4
Tracy	Smith	220	4
Gerald	Butters	210	6
Burton	Tracy	180	7
Charles	Owen	170	8
Janice	Joplette	160	9
Anne	Baker	150	10

Rank members by (rounded) hours used

QUESTION

Produce a list of members (including guests), along with the number of hours they've booked in facilities, rounded to the nearest ten hours.

Rank them by this rounded figure, producing output of first name, surname, rounded hours, rank. Sort by rank, surname, and first name.

EXPECTED RESULT

firstname	surname	hours	rank
GUEST	GUEST	1200	1
Darren	Smith	340	2
Tim	Rownam	330	3
Tim	Boothe	220	4
Tracy	Smith	220	4
Gerald	Butters	210	6
Burton	Tracy	180	7
Charles	Owen	170	8
Janice	Joplette	160	9
Anne	Baker	150	10

Find the top three revenue generating facilities

QUESTION

Produce a list of the top three revenue generating facilities (including ties).

Output facility name and rank, sorted by rank and facility name.

EXPECTED RESULT

name	rank
Massage Room 1	1
Massage Room 2	2
Tennis Court 2	3

Classify facilities by value

QUESTION

Classify facilities into equally sized groups of high, average, and low based on their revenue.

Order by classification and facility name.

EXPECTED RESULT

name	revenue
Massage Room 1	high
Massage Room 2	high
Tennis Court 2	high
Badminton Court	average
Squash Court	average
Tennis Court 1	average
Pool Table	low
Snooker Table	low
Table Tennis	low

Calculate the payback time for each facility

QUESTION

Based on the 3 complete months of data so far, calculate the amount of time each facility will take to repay its cost of ownership.

Remember to take into account ongoing monthly maintenance. Output facility name and payback time in months, order by facility name.

Don't worry about differences in month lengths, we're only looking for a rough value here!

EXPECTED RESULT

name	months
Badminton Court	6.8317677198975235
Massage Room 1	0.18885741265344664778
Massage Room 2	1.7621145374449339
Pool Table	5.3333333333333333
Snooker Table	6.9230769230769231
Squash Court	1.1339582703356516
Table Tennis	6.4000000000000000
Tennis Court 1	2.2624434389140271
Tennis Court 2	1.7505470459518600

Calculate the rolling average of total revenue

QUESTION

For each day in August 2012, calculate a rolling average of total revenue over the previous 15 days.

Output should contain date and revenue columns, sorted by the date.

Remember to account for the possibility of a day having zero revenue.

This one's a bit tough, so don't be afraid to check out the hint!

EXPECTED RESULT

date	revenue
2012-08-01	1126.8333333333333
2012-08-02	1153.0000000000000000
2012-08-03	1162.9000000000000000
2012-08-04	1177.3666666666666667
2012-08-05	1160.9333333333333333
2012-08-06	1185.4000000000000000
2012-08-07	1182.8666666666666667
2012-08-08	1172.6000000000000000
2012-08-09	1152.4666666666666667
2012-08-10	1175.0333333333333333