#### **SQL - Lab 2: EdX Database Exercise**

## Part 1: An exploration of 6.00x

How many enrollees are in the class with course\_id MITx/6.00x/2012\_Fall? Assumed enrolled is the same as "registered."

```
SELECT course_id, count (*)
FROM RECORDS
WHERE course_id = 'MITx/6.00x/2012_Fall'
AND registered = 1;
```

	course_id	count (*)
1	MITx/6.00x/2012_Fall	6804

There are 6,804 enrollees in this particular class in fall.

That same course was also taught again the following term with course\_id "MITx/6.00x/2013\_Spring". Did enrollment go up or down?

```
SELECT course_id, count (*)
FROM RECORDS
WHERE course_id = 'MITx/6.00x/2013_Spring'
AND registered = 1;
```

course_id	count (*)
1 MITx/6.00x/2013_Spring	5775

Looks like enrollment has went down, from 6804 to 5775

What was the average grade in each term of 6.00x?

```
SELECT course_id, avg(grade)
FROM RECORDS
WHERE course_id = 'MITx/6.00x/2012_Fall'
AND registered = 1;
```

course_id	avg(grade)
1 MITx/6.00x/2012_Fall	0.0443974132863017

SELECT course\_id, avg(grade)
FROM RECORDS
WHERE course\_id = 'MITx/6.00x/2013\_Spring'
AND registered = 1;

course_id		avg(grade)	
ŀ	1 MITx/6.00x/2013_Spring	0.028195670995671	

The average grade for the Fall term is 0.044, and 0.028 in the Spring term.

Uh oh, those seem too low. Remove all of the zeroes for people who haven't taken any tests and calculate the average from the remaining data for both terms of 6.00x?

```
SELECT course_id, avg(grade)
FROM RECORDS
WHERE course_id = 'MITx/6.00x/2012_Fall'
AND registered = 1
AND NOT grade = 0;
```

course_id	avg(grade)		
1 MITx/6.00x/2012_Fall	0.199261213720314		

```
SELECT course_id, avg(grade)
FROM RECORDS
WHERE course_id = 'MITx/6.00x/2013_Spring'
AND registered = 1
AND NOT grade = 0;
```

course_id	avg(grade)
1 MITx/6.00x/2013_Spring	0.0861078794288736

The normalized average grade for the Fall term is 0.19, and 0.086 in the Spring term.

What was the total number of enrollments for 6.00x over both terms?

```
SELECT count (*)
FROM RECORDS
WHERE course_id = 'MITx/6.00x/2012_Fall'
OR course_id = 'MITx/6.00x/2013_Spring';
```

```
count (*)
1 12579
```

How many people were enrolled (registered) in both terms of 6.00x?

```
SELECT count (*)
FROM RECORDS
WHERE (course_id = 'MITx/6.00x/2012_Fall'
OR course_id = 'MITx/6.00x/2013_Spring')
AND registered = 1;

count (*)
1 12579
```

#### Part 2: A look at all courses

Let's take this one step further, let's list the number of enrollees in every course.

SELECT DISTINCT(course\_id), Count (\*)
FROM RECORDS
GROUP BY course\_id;

	course_id	Count (*)
1	HarvardX/CB22x/2013_Spring	2917
2	HarvardX/CS50x/2012	16926
3	HarvardX/ER22x/2013_Spring	5686
4	HarvardX/PH207x/2012_Fall	4199
5	HarvardX/PH278x/2013_Spring	4044
6	MITx/14.73x/2013_Spring	2715
7	MITx/2.01x/2013_Spring	573
8	MITx/3.091x/2012_Fall	1398
9	MITx/3.091x/2013_Spring	582
10	MITx/6.002x/2012_Fall	4053
11	MITx/6.002x/2013_Spring	2196
12	MITx/6.00x/2012_Fall	6804
13	MITx/6.00x/2013_Spring	5775
14	MITx/7.00x/2013_Spring	2032
15	MITx/8.02x/2013_Spring	3045
16	MITx/8.MReV/2013_Summer	937

Make the returned result nicer by changing the header from COUNT(\*) to enrollees.

# SELECT DISTINCT(course\_id), Count (\*) as enrollees FROM RECORDS GROUP BY course\_id;

	course_id	enrollees
1	HarvardX/CB22x/2013_Spring	2917
2	HarvardX/CS50x/2012	16926
3	HarvardX/ER22x/2013_Spring	5686
4	HarvardX/PH207x/2012_Fall	4199
5	HarvardX/PH278x/2013_Spring	4044
6	MITx/14.73x/2013_Spring	2715
7	MITx/2.01x/2013_Spring	573
8	MITx/3.091x/2012_Fall	1398
9	MITx/3.091x/2013_Spring	582
10	MITx/6.002x/2012_Fall	4053
11	MITx/6.002x/2013_Spring	2196
12	MITx/6.00x/2012_Fall	6804
13	MITx/6.00x/2013_Spring	5775
14	MITx/7.00x/2013_Spring	2032
15	MITx/8.02x/2013_Spring	3045
16	MITx/8.MReV/2013_Summer	937

Order the previous results to show the number of enrollees in descending order. Which class had the most enrollees?

SELECT DISTINCT(course\_id), Count (\*) as enrollees FROM RECORDS
GROUP BY course\_id
ORDER BY enrollees DESC;

	course_id	enrollees
1	HarvardX/CS50x/2012	16926
2	MITx/6.00x/2012_Fall	6804
3	MITx/6.00x/2013_Spring	5775
4	HarvardX/ER22x/2013_Spring	5686
5	HarvardX/PH207x/2012_Fall	4199
6	MITx/6.002x/2012_Fall	4053
7	HarvardX/PH278x/2013_Spring	4044
8	MITx/8.02x/2013_Spring	3045
9	HarvardX/CB22x/2013_Spring	2917
10	MITx/14.73x/2013_Spring	2715
11	MITx/6.002x/2013_Spring	2196
12	MITx/7.00x/2013_Spring	2032
13	MITx/3.091x/2012_Fall	1398
14	MITx/8.MReV/2013_Summer	937
15	MITx/3.091x/2013_Spring	582
16	MITx/2.01x/2013_Spring	573

The CS50x course from Harvard that took place in 2012 had the most enrollees.

# Part 3: A deeper look at all of the courses

Create the same list of enrollees in descending order, but include only the course long title and the counts of enrollees in the returned result. Note this will group together courses with the same name, which were taught in different terms.

SELECT DISTINCT(Course\_Long\_Title), Count (\*) as enrollees FROM RECORDS
GROUP BY Course\_Long\_Title
ORDER BY enrollees DESC;

	Course_Long_Title	enrollees
1	Introduction to Computer Science 1	16926
2	Introduction to Computer Science and Programming	12579
3	Circuits and Electronics	6249
4	Justice	5686
5	Health in Numbers: Quantitative Methods in Clinical & Public Health Research	4199
6	Human Health and Global Environmental Change	4044
7	Electricity and Magnetism	3045
8	The Ancient Greek Hero	2917
9	The Challenges of Global Poverty	2715
10	Introduction to Biology _ The Secret of Life	2032
11	Introduction to Solid State Chemistry	1980
12	Mechanics Review	937
13	Elements of Structures	573

Now we need to find out which are the courses that had at least 4,000 Enrollees? You may do this by course id.

SELECT course\_id, Count(\*) as enrollees FROM RECORDS GROUP BY course\_id HAVING enrollees > 4000;

	course_id	enrollees
1	HarvardX/CS50x/2012	16926
2	HarvardX/ER22x/2013_Spring	5686
3	HarvardX/PH207x/2012_Fall	4199
4	HarvardX/PH278x/2013_Spring	4044
5	MITx/6.002x/2012_Fall	4053
6	MITx/6.00x/2012_Fall	6804
7	MITx/6.00x/2013_Spring	5775

## Part 4: Course engagement

Now let's put together some statistics for the engagement for each course. How many people are registered, have viewed, have explored, and have become certified for each course by course\_id? Rename the columns.

SELECT course\_id, sum(registered) as registered, sum(viewed) as viewed, sum(explored) as explored, sum(certified) as certified
FROM Course\_users
GROUP BY course\_id;

	course_id	registered	viewd	explored	certified
1	HarvardX/CB22x/	2917	1604	57	45
2	HarvardX/CS50x/	16926	10654	1142	141
3	HarvardX/ER22x/	5686	3150	347	227
4	HarvardX/PH207x	4199	2443	467	193
5	HarvardX/PH278x	4044	1508	121	77
6	MITx/14.73x/	2715	1604	294	208
7	MITx/2.01x/	573	384	59	25
8	MITx/3.091x/	1398	705	86	53
9	MITx/3.091x/	582	560	17	14
10	MITx/6.002x/	4053	2606	280	170
11	MITx/6.002x/	2196	1057	102	73
12	MITx/6.00x/	6804	4207	464	280
13	MITx/6.00x/	5775	5442	271	128
14	MITx/7.00x/	2032	1237	148	87
15	MITx/8.02x/	3045	2060	175	79
16	MITx/8.MReV/	937	636	36	26

Challenge: Redo the above question using course long title instead of course\_id. Note this will group together courses with the same name, which were taught in different terms.

	Course_Long_Title	registered	viewed	explored	certified
1	Circuits and Electronics	6249	3663	382	243
2	Electricity and Magnetism	3045	2060	175	79
3	Elements of Structures	573	384	59	25
4	Health in Numbers: Quantitative Methods in Clinical & Public Health Research	4199	2443	467	193
5	Human Health and Global Environmental Change	4044	1508	121	77
6	Introduction to Biology _ The Secret of Life	2032	1237	148	87
7	Introduction to Computer Science 1	16926	10654	1142	141
8	Introduction to Computer Science and Programming	12579	9649	735	408
9	Introduction to Solid State Chemistry	1980	1265	103	67
10	Justice	5686	3150	347	227
11	Mechanics Review	937	636	36	26
12	The Ancient Greek Hero	2917	1604	57	45
13	The Challenges of Global Poverty	2715	1604	294	208

What fraction of the users view, explore, or certify in the content in each course once they have registered? Rename the columns. For a challenge, do this again with the long course title, noting that this will group courses with the same name across terms.

	course_id	registered	viewed	view_frac	explored	exp_frac	certified	cert_frac
1	HarvardX/CB22x/2013_Spring	2917	1604	0.55	57	0.02	45	0.02
2	HarvardX/CS50x/2012	16926	10654	0.63	1142	0.07	141	0.01
3	HarvardX/ER22x/2013_Spring	5686	3150	0.55	347	0.06	227	0.04
4	HarvardX/PH207x/2012_Fall	4199	2443	0.58	467	0.11	193	0.05
5	HarvardX/PH278x/2013_Spring	4044	1508	0.37	121	0.03	77	0.02
6	MITx/14.73x/2013_Spring	2715	1604	0.59	294	0.11	208	0.08
7	MITx/2.01x/2013_Spring	573	384	0.67	59	0.1	25	0.04
8	MITx/3.091x/2012_Fall	1398	705	0.5	86	0.06	53	0.04
9	MITx/3.091x/2013_Spring	582	560	0.96	17	0.03	14	0.02
10	MITx/6.002x/2012_Fall	4053	2606	0.64	280	0.07	170	0.04
11	MITx/6.002x/2013_Spring	2196	1057	0.48	102	0.05	73	0.03
12	MITx/6.00x/2012_Fall	6804	4207	0.62	464	0.07	280	0.04
13	MITx/6.00x/2013_Spring	5775	5442	0.94	271	0.05	128	0.02
14	MITx/7.00x/2013_Spring	2032	1237	0.61	148	0.07	87	0.04
15	MITx/8.02x/2013_Spring	3045	2060	0.68	175	0.06	79	0.03
16	MITx/8.MReV/2013_Summer	937	636	0.68	36	0.04	26	0.03

	Course_Long_Title	registered	viewed	view_frac	explored	exp_frac	certified	cert_frac
1	Circuits and Electronics	6249	3663	0.59	382	0.06	243	0.04
2	Electricity and Magnetism	3045	2060	0.68	175	0.06	79	0.03
3	Elements of Structures	573	384	0.67	59	0.1	25	0.04
4	Health in Numbers: Quantitative Methods in Clinical &	4199	2443	0.58	467	0.11	193	0.05
5	Human Health and Global Environmental Change	4044	1508	0.37	121	0.03	77	0.02
6	Introduction to Biology _ The Secret of Life	2032	1237	0.61	148	0.07	87	0.04
7	Introduction to Computer Science 1	16926	10654	0.63	1142	0.07	141	0.01
8	Introduction to Computer Science and Programming	12579	9649	0.77	735	0.06	408	0.03
9	Introduction to Solid State Chemistry	1980	1265	0.64	103	0.05	67	0.03
10	Justice	5686	3150	0.55	347	0.06	227	0.04
11	Mechanics Review	937	636	0.68	36	0.04	26	0.03
12	The Ancient Greek Hero	2917	1604	0.55	57	0.02	45	0.02
13	The Challenges of Global Poverty	2715	1604	0.59	294	0.11	208	0.08

## Part 5: More challenging questions

Find the list of courses that are hosted at HarvardX and have more than 4000 enrollees?

```
SELECT *, Count(*) as enrollees
FROM(
SELECT *
FROM(
SELECT Course.course_id, Course.Course_Long_Title
FROM Course
JOIN course_users
ON Course.course_id = course_users.course_id)
WHERE course_id LIKE 'Harvard%')
GROUP BY course_id
HAVING enrollees > 4000
```

	course_id	Course_Long_Title	enrollees
1	HarvardX/CS50x/2012	Introduction to Computer Science 1	16926
2	HarvardX/ER22x/2013_Spring	Justice	5686
3	HarvardX/PH207x/2012_Fall	Health in Numbers: Quantitative Methods in Clinical & Public Health	4199
4	HarvardX/PH278x/2013_Spring	Human Health and Global Environmental Change	4044

How many users who have registered more than 3 courses?

```
SELECT *, count(*) as course_enrolled
FROM(
SELECT Course.course_id, Course_users.userid_DI
FROM Course
JOIN course_users
ON Course.course_id = course_users.course_id)
GROUP BY userid_DI
HAVING course_enrolled > 3;
```

	course_id	userid_DI	course_enrolled
1	HarvardX/PH278x/2013_Spring	MHxPC130034504	4
2	MITx/14.73x/2013_Spring	MHxPC130061330	4
3	MITx/3.091x/2012_Fall	MHxPC130075396	4
4	HarvardX/CS50x/2012	MHxPC130103172	4
5	HarvardX/CS50x/2012	MHxPC130126780	5
6	HarvardX/CS50x/2012	MHxPC130292057	5
7	HarvardX/PH207x/2012_Fall	MHxPC130414295	4
8	MITx/2.01x/2013_Spring	MHxPC130464954	5
9	HarvardX/PH278x/2013_Spring	MHxPC130509722	4
10	MITx/6.002x/2013_Spring	MHxPC130555501	4
11	HarvardX/CB22x/2013_Spring	MHxPC130590508	4

```
SELECT count(*)
FROM(
SELECT *, count(*) as course_enrolled
FROM(
SELECT Course.course_id, Course_users.userid_DI
FROM Course
JOIN course_users
ON Course.course_id = course_users.course_id)
GROUP BY userid_DI
HAVING course_enrolled > 3);
```



So there are 11 users that are enrolled in more than 3 courses

How many users are there by country? Order these alphabetically by country.

SELECT Country, Count (\*) as user\_num FROM Users GROUP BY Country ORDER BY Country ASC;

	Country	user_num
1	Australia	610
2	Bangladesh	303
3	Brazil	1765
4	Canada	1233
5	China	513
6	Colombia	443
7	Egypt	874
8	France	441
9	Germany	787
10	Greece	526
11	India	8566
12	Indonesia	331
13	Japan	217
14	Mexico	500
15	Morocco	394
16	Nigeria	709
17	Other Africa	2351
18	Other East Asia	611
19	Other Europe	3800
20	Other Middle East	1612
21	Other North &	432
22	Other Oceania	38
23	Other South	991
24	Other South Asia	1260
25	Pakistan	1063
26	Philippines	499
27	Poland	492
28	Portugal	197
29	Russian Federation	982
30	Spain	1020
31	Ukraine	419
32	United Kingdom	2084
33	United States	17491
34	Unknown/Other	7983

What is the average grade by country for users who have become certified in any course? Order by average grade, descending.

```
SELECT Country, count (*) as certificated_user, round(avg (grade),3) as avg_grade
FROM(
SELECT Users.Country, Users.userid_DI, course_users.registered, course_users.grade
FROM Users
JOIN course_users
ON Users.userid_DI = course_users.userid_DI
WHERE certified = 1
)
GROUP BY Country
ORDER BY avg_grade DESC
```

	Country	certificated_user	avg_grade
1	Morocco	3	0.927
2	France	13	0.894
3	Ukraine	20	0.881
4	Australia	29	0.881
5	Other Africa	50	0.878
6	Egypt	16	0.868
7	Other North &	19	0.858
8	Spain	90	0.856
9	Russian Federation	68	0.854
10	Other Europe	196	0.853
11	Mexico	15	0.852
12	Unknown/Other	11	0.849
13	Germany	51	0.848
14	Poland	44	0.84
15	United Kingdom	91	0.839

16	India	347	0.836
17	Canada	28	0.835
18	Other South	35	0.831
19	Portugal	8	0.83
20	Nigeria	12	0.828
21	Pakistan	17	0.821
22	Japan	4	0.82
23	Greece	26	0.82
24	Other Middle East	27	0.818
25	United States	462	0.817
26	Other South Asia	38	0.815
27	China	8	0.815
28	Colombia	18	0.812
29	Philippines	5	0.802
30	Indonesia	12	0.802
31	Brazil	46	0.796
32	Other East Asia	14	0.787
33	Bangladesh	3	0.783

Which country has the highest average grade for people certified in a course? Which country has the lowest average grade for people certified in a course?

```
SELECT Country, count (*) as certificated_user, round(avg (grade),3) as avg_grade
FROM(
SELECT Users.Country, Users.userid_DI, course_users.registered, course_users.grade
FROM Users
JOIN course_users
ON Users.userid_DI = course_users.userid_DI
WHERE certified = 1
)
GROUP BY Country
ORDER BY avg_grade DESC
LIMIT 1;
```

	Country	certificated_user	avg_grade
1	Morocco	3	0.927

SELECT Country, count (\*) as certificated\_user, round(avg (grade),3) as avg\_grade

```
FROM(
SELECT Users.Country, Users.userid DI, course users.registered,
course_users.grade
FROM Users
JOIN course users
ON Users.userid DI = course users.userid DI
WHERE certified = 1
GROUP BY Country
ORDER BY avg grade ASC
LIMIT 1;
              certificated user
    Country
                               avg grade
                             3
                                     0.783
 1 Bangladesh
```

## Part 6: Harvard and MIT, course grades

What is the average grade by country for users who have become certified in any *HARVARD* course? Order by average grade, descending.

	Country	certificated_user	avg_grade
1	Ukraine	5	1.0
2	Portugal	4	0.958
3	Other North & Central Amer	5	0.954
4	Pakistan	7	0.937
5	Other Africa	37	0.915
6	Other Europe	61	0.908
7	Egypt	9	0.906
8	Australia	14	0.904
9	United Kingdom	41	0.899
10	France	7	0.899
11	Other Middle East/Central Asia	12	0.898
12	Russian Federation	13	0.892
13	India	117	0.887
14	Other South America	8	0.883
15	Canada	14	0.877
16	Unknown/Other	3	0.87
17	Spain	26	0.868
18	Poland	5	0.866
19	Germany	29	0.865
20	Indonesia	5	0.86
21	Mexico	6	0.857
22	United States	163	0.849
23	Philippines	4	0.843
24	China	7	0.839
25	Other South Asia	23	0.837
26	Colombia	5	0.834
27	Japan	3	0.833
28	Nigeria	10	0.826
29	Brazil	19	0.822
30	Other East Asia	4	0.815
31	Greece	16	0.766
32	Bangladesh	1	0.61

Which country has the highest average grade for people certified in a HARVARD course? Which country has the lowest average grade for people certified in a HARVARD course? SELECT Country, count (\*) as certificated\_user, round(avg (grade),3) as avg\_grade

FROM(

SELECT Users.Country, Users.userid\_DI, course\_users.certified, course\_users.grade,
Course.course\_id

**FROM Course** 

LEFT JOIN course users

ON (Course\_id = course\_users.course\_id)

**LEFT JOIN Users** 

ON (course\_users.userid\_DI = Users.userid\_DI)

WHERE certified = 1

AND course\_users.course\_id like 'Harvard%')

**GROUP BY Country** 

ORDER BY avg\_grade DESC

LIMIT 1;

	Country	certificated_user	avg_grade
1	Ukraine	5	1.0

SELECT Country, count (\*) as certificated\_user, round(avg (grade),3) as avg\_grade

FROM(

SELECT Users.Country, Users.userid\_DI, course\_users.certified, course\_users.grade, Course\_course\_id

FROM Course

LEFT JOIN course\_users

ON (Course\_id = course\_users.course\_id)

**LEFT JOIN Users** 

ON (course\_users.userid\_DI = Users.userid\_DI)

WHERE certified = 1

AND course\_users.course\_id like 'Harvard%')

**GROUP BY Country** 

ORDER BY avq\_grade ASC

LIMIT 1;

	Country	certificated_user	avg_grade
1	Bangladesh	1	0.61

#### Repeat query for MIT.

SELECT Country, count (\*) as certificated\_user, round(avg (grade),3) as avg\_grade

FROM(

SELECT Users.Country, Users.userid\_DI, course\_users.certified, course\_users.grade, Course.course\_id

**FROM Users** 

JOIN course users

ON Users.userid\_DI = course\_users.userid\_DI

JOIN Course
ON course\_users.course\_id = Course.course\_id
WHERE certified = 1
AND Course.course\_id like 'MIT%')
GROUP BY Country
ORDER BY avg\_grade DESC;

	Country	certificated_user	avg_grade
1	Morocco	3	0.927
フー	Greece	10	0.906
3	France	6	0.888
4	Bangladesh	2	0.87
5	Australia	15	0.859
h	Spain	64	0.851
/	Mexico	9	0.849
X	Russian Federation	55	0.845
ч	Unknown/Other	8	0.841
10	Ukraine	15	0.841
	Poland	39	0.836
	Nigeria	2	0.835
13	Other Europe	135	0.828
14	Germany	22	0.825
15	Other North & Central Amer. and Caribbean	14	0.824
	Egypt	7	0.82
	Other South America	27	0.816
18	India	230	0.81
19	Colombia	13	0.804
20	United States	299	0.8
	Canada	14	0.794
77	United Kingdom	50	0.79
73	Other South Asia	15	0.78
74	Japan	1	0.78
	Brazil	27	0.777
	Other East Asia	10	0.776
77	Other Africa	13	0.772
	Indonesia	7	0.76
79	Other Middle East/Central Asia	15	0.754
30	Pakistan	10	0.74
	Portugal	4	0.703
	China	1	0.65
33	Philippines	1	0.64

SELECT Country, count (\*) as certificated\_user, round(avg (grade),3) as avg\_grade FROM(

SELECT Users.Country, Users.userid\_DI, course\_users.certified, course\_users.grade, Course.course\_id

**FROM Course** 

LEFT JOIN course\_users

ON (Course.course\_id = course\_users.course\_id)

**LEFT JOIN Users** 

ON (course\_users.userid\_DI = Users.userid\_DI)

WHERE certified = 1

AND course\_users.course\_id like 'MIT%')

GROUP BY Country
ORDER BY avg\_grade DESC
LIMIT 1;

	Country	certificated_user	avg_grade
1	Morocco	3	0.927

SELECT Country, count (\*) as certificated\_user, round(avg (grade),3) as avg\_grade FROM(

SELECT Users.Country, Users.userid\_DI, course\_users.certified, course\_users.grade, Course.course\_id

**FROM Course** 

LEFT JOIN course\_users

ON (Course.course\_id = course\_users.course\_id)

**LEFT JOIN Users** 

ON (course\_users.userid\_DI = Users.userid\_DI)

WHERE certified = 1

AND course\_users.course\_id like 'MIT%')

**GROUP BY Country** 

ORDER BY avg\_grade ASC

LIMIT 1;

	Country	certificated_user	avg_grade
1	Philippines	1	0.64