

Homework 3

SUBJECT: Week 3

FROM: Zhonghan Xie

TO: Raj Kumar

Hi Raj,

THanks for your email. I've looked into the **taxdata** database to know more about the non-profits' tax data and answer your questions.

Part 1:

1. In the year 2014, there were 247 Ann Arbor companies in the database.

```
[mysql> SELECT COUNT(*) AS ent_in_aa FROM taxdata WHERE city='ANN ARBOR' AND year=2014;
+-----+
| ent_in_aa |
+-----+
|      247 |
+-----+
1 row in set (0.53 sec)
```

2. In the year 2014, there were 20 companies making more than ten billion dollars in revenue. The companies are listed below:

```
[mysql> SELECT name, revenue, year FROM taxdata WHERE revenue>10000000000 AND year=2014 ORDER BY revenue DESC;
+-----+-----+-----+
| name                                     | revenue | year |
+-----+-----+-----+
| UAW RETIREE MEDICAL BENEFITS TRUST      | 118834955469 | 2014 |
| President and Fellows of Harvard College | 90325789860  | 2014 |
| CENTRAL STATES SE & SW AREAS HEALTH & WELFARE F | 90230569901  | 2014 |
| KAISER FOUNDATION HEALTH PLAN INC       | 66501905671  | 2014 |
| Howard Hughes Medical Institute         | 49190053351  | 2014 |
| THE BOARD OF TRUSTEES OF THE LELAND STANFORDJUNIOR UNIVERSITY | 33572755723  | 2014 |
| KAISER FOUNDATION HOSPITALS             | 30499444071  | 2014 |
| Massachusetts Institute of Technology   | 26659611000  | 2014 |
| Thrivent Financial for Lutherans         | 25666577042  | 2014 |
| STATE EMPLOYEES' CREDIT UNION           | 16098699427  | 2014 |
| Trustees of the University of Pennsylvania | 15876299641  | 2014 |
| TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA | 15876299641  | 2014 |
| IHC HEALTH SERVICES INC                 | 15110938763  | 2014 |
| DUKE UNIVERSITY                        | 13001041449  | 2014 |
| Banner Health                           | 12799110462  | 2014 |
| DIGNITY HEALTH                          | 11279915619  | 2014 |
| YALE UNIVERSITY                         | 11199355037  | 2014 |
| Partners HealthCare System Inc & AffiliatesGroup Return | 11123194520  | 2014 |
| UPMC GROUP                             | 10923272930  | 2014 |
| Cornell University                      | 10721163381  | 2014 |
+-----+-----+-----+
20 rows in set (0.50 sec)
```

3. Over the years, there were 24 companies making more than ten billion dollars in revenue. The companies are listed in the terminal output below. It seems that most of the companies making more

than ten billion dollars in revenue are in the year 2014.

```
[mysql> SELECT name, revenue, year FROM taxdata WHERE revenue>10000000000 ORDER BY revenue DESC;
```

name	revenue	year
UAW RETIREE MEDICAL BENEFITS TRUST	118834955469	2014
President and Fellows of Harvard College	90325789860	2014
CENTRAL STATES SE & SW AREAS HEALTH & WELFARE F	90230569901	2014
KAISER FOUNDATION HEALTH PLAN INC	75534372412	2012
KAISER FOUNDATION HEALTH PLAN INC	66889601170	2013
KAISER FOUNDATION HEALTH PLAN INC	66501905671	2014
Howard Hughes Medical Institute	49190053351	2014
THE BOARD OF TRUSTEES OF THE LELAND STANFORDJUNIOR UNIVERSITY	33572755723	2014
KAISER FOUNDATION HOSPITALS	30499444071	2014
Massachusetts Institute of Technology	26659611000	2014
Thrivent Financial for Lutherans	25666577042	2014
KAISER FOUNDATION HOSPITALS	24923206152	2012
STATE EMPLOYEES' CREDIT UNION	16098699427	2014
Trustees of the University of Pennsylvania	15876299641	2014
TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA	15876299641	2014
IHC HEALTH SERVICES INC	15110938763	2014
DUKE UNIVERSITY	13001041449	2014
Banner Health	12799110462	2014
DIGNITY HEALTH	11279915619	2014
YALE UNIVERSITY	11199355037	2014
Partners HealthCare System Inc & AffiliatesGroup Return	11123194520	2014
WORKERS' COMPENSATION REINSURANCE ASSOCIATION	10931957168	2015
UPMC GROUP	10923272930	2014
Cornell University	10721163381	2014

24 rows in set (0.52 sec)

Part 2:

1. The top 20 companies with the most expenses in the year 2013 are listed below.

```
[mysql> SELECT name, year, expenses FROM taxdata WHERE year=2013 ORDER BY expenses DESC LIMIT 20;
```

name	year	expenses
KAISER FOUNDATION HEALTH PLAN INC	2013	41982288055
UNIVERSITY OF ROCHESTER	2013	2776536619
ST FRANCIS HOSPITAL AND MEDICAL CENTER	2013	713077460
WELLSTAR HEALTH SYSTEM INC	2013	660400156
SKY LAKES MEDICAL CENTER	2013	488333644
NAPLES COMMUNITY HOSPITAL INC	2013	426168844
YAKIMA VALLEY MEMORIAL HOSPITAL	2013	401985865
EMBRY-RIDDLE AERONAUTICAL UNIVERSITY INC	2013	374154000
ST MARY'S HOSPITAL & MEDICAL CENTERINC	2013	372041256
St Barnabas Hospital	2013	349321637
HOSPITAL COMMITTEE FOR THE LIVERMOREPLEASANTON AREA	2013	341257233
VALLEY PRESBYTERIAN HOSPITAL	2013	273789797
WINCHESTER HOSPITAL	2013	264605951
TEAMSTERS WESTERN REGION AND NEW JERSEYHEALTH CARE FUND	2013	231059337
OROVILLE HOSPITAL	2013	200242937
ST JOSEPH HOSPITAL OF EUREKA	2013	196999313
SWARTHMORE COLLEGE	2013	168742729
ST FRANCIS HOSPITAL - POUGHKEEPSIE	2013	164506124
Vail Clinic Inc	2013	159617575
TUSKEGEE UNIVERSITY	2013	152572824

20 rows in set (0.44 sec)

2. There were 88153 companies with a revenue between \$1 and \$100,000 and expenses between \$10,000 and \$200,000 in our records. I randomly selected 20 companies from the list. Their EINs, names, cities and their revenues and expenses are listed below.

```
[mysql> SELECT COUNT(*) FROM taxdata WHERE ( revenue BETWEEN 1 AND 100000) AND (expenses BETWEEN 10000 AND 200000);
```

COUNT(*)
88153

```
1 row in set (0.39 sec)
```

```
[mysql> SELECT ein, name, city, revenue, expenses FROM taxdata WHERE ( revenue BETWEEN 1 AND 100000) AND (expenses BETWEEN 10000 AND 200000) ORDER BY RAND() LIMIT 20;
```

ein	name	city	revenue	expenses
731445844	Starbase Oklahoma Inc	Tulsa	59875	47031
222956306	SHREWSBURY HOSE COMPANY NO 1 INC	SHREWSBURY	66502	52407
251587927	MARS ROD & GUN CLUB INC	MARS	16169	40216
946174137	SOROPTIMIST INTL OF TWAIN HARTE	TWAIN HARTE	48001	44773
954696668	CHESTER HOUSING FOUNDATION	Pasadena	80750	67134
421440222	PLEASANT HILL SOCCER CLUB	DES MOINES	63142	64576
916185837	Kirkman House Museum	Walla Walla	44427	42972
231542761	SCHUYLKILL COUNTY BAR ASSOCIATION	POTTSVILLE	93436	78096
231996042	NATIONAL SORORITY OF PHI DELTA KAPPA ZETA CHAPTER	PHILADELPHIA	39934	34817
746107046	Comfort Chamber of Commerce	Comfort	67226	64456
953188134	PORTUGUESE HISTORICAL CENTER	SAN DIEGO	59608	53486
341517534	COMMON WEALTH REVOLVING LOAN FUND	KENT	37283	165945
203801172	JEROME LOHEZ SEPTEMBER 11 SCHOLARSHIPFOUNDATION	NEW YORK	10264	22234
900249104	Hillsborough Raiders Touchdown Association Inc	Hillsborough	73268	62872
273941723	LOS ANGELES POLICE EMERALD SOCIETYPipes AND DRUMS INC	LOS ANGELES	53093	38438
202834496	Friends of the Auberry Library	Auberry	35249	94736
311492157	BRAYDN & BRHET SUPPORT FOUNDATION	MURRAY	75987	74868
263555871	WOMANS CLUB OF DELRAY BEACH INC	DELRAY BEACH	15363	14811
592947435	CLEARWATER - ST PETERSBURG - TAMPAGOLDEN TRIANGLE ASSOCIATION INC	ST PETERSBURG	60662	55871
371562868	MINOOKA LACROSSE CORP	MINOOKA	66339	55770

```
20 rows in set (0.67 sec)
```

Part 3:

1. There are 378 companies with "toy" mentioned in their purposes.

```
[mysql> SELECT COUNT(*) FROM taxdata WHERE purpose LIKE '%toy%';
```

COUNT(*)
378

```
1 row in set (2.54 sec)
```

2. There are 2794 companies with "smith" mentioned in their **ptname** field and with a positive revenue in our records.

```
[mysql> SELECT COUNT(*) FROM taxdata WHERE ptname LIKE '%smith%' AND revenue>0;
```

COUNT(*)
2794

```
1 row in set (0.60 sec)
```

Part 4:

1. I pulled the name and the length of the names of 50 randomly selected companies with **ptid=P01345770** in the **taxdata** table.

```
[mysql> SELECT name, LENGTH(name) FROM taxdata WHERE ptid='P01345770' ORDER BY RAND() LIMIT 50;
```

name	LENGTH(name)
The Jordan-Evans Family Foundation	34
The Red Cabin Foundation Trust	30
The Andrew and Barbara Bangser CharitableFoundation Inc	55
The Scrivan Family Foundation	29
The Obrzut - Ling Foundation Inc	32
The Dowling & Partners Charitable FoundationInc	47
Jerome A Fink Foundation	24
The Bruce G and Mary A Robert FamilyFoundation Inc	50
Amicus Amico Foundation	23
Chauncey Foundation	19
SALLY AND DICK ROBERTS COYOTE FOUNDATION	40
The Douglas and Elizabeth Durand FamilyFoundation Inc	53
The Roy & Betty Wilde Foundation	32
Bharat Family Foundation	24
Soran Foundation	16
Edris Charitable Foundation	27
The Richard E Frain Jr CharitableFoundation Inc	47
The HOPE Charitable Foundation	30
The Steven C Calicchio Foundation	33
The Doherty Foundation	22
The David E Maguire Foundation Inc	34
D&P Roberts Family Foundation	29
Rosen Family Foundation	23
The Justin and Michelle Hughes Foundation	41
Lighten Family Foundation	25
CVW Family Foundation	21
The Ho Hing Dai Family Foundation Inc	37
Everything is Possible Foundation	33
Eternal Perspectives Foundation	31
Michael and Ina Korek Foundation Trust	38
Bevan Daddino Foundation Inc	28
The Steve Tirado Family Foundation	34
The Caswell Jin Foundation	26
The Burt's Bees Greater Good Foundation	39
The H Private Foundation	24
John and Diane Scelfo Charitable Foundation	43
Felvis Foundation	17
Joseph & Ruth Scodari Family Foundation	39
Richard Lumsden Foundation	26
Neill M Timmons Foundation	26
Replogle Family Foundation	26
Kerry and Laura Bradley Foundation	34
The Paul Scammell Foundation	28
The Schoenholtz Family Foundation	33
The Fries Foundation	20
The Phyllis M Coors Foundation	30
The Dr Michael J Welch Foundation	33
Barcus King Family Foundation	29
The Timothy and Judith Smith Foundation	39
The WJK Charitable Foundation Inc	33

```
50 rows in set (0.49 sec)
```

2. There are 917 companies with a **purpose** field containing less than 10 characters in our records.

```
[mysql> SELECT COUNT(*) FROM taxdata WHERE LENGTH(purpose)>0 AND LENGTH(purpose)<10;
```

COUNT(*)
917

```
1 row in set (0.63 sec)
```

Part 5:

1. There are 52,560 employees that were hired in the years of 1990, 1994 and 1995. The shortest **WHERE** clause is **WHERE year(hire_date) IN (1990, 1994, 1995)**.

```
[mysql> SELECT COUNT(*) FROM employees WHERE YEAR(hire_date) IN (1994, 1995, 1990);
```

COUNT(*)
52560

```
1 row in set (0.16 sec)
```

2. There were 2,796 Senior Engineers in the company on 1986-06-26. The query is `SELECT count(*) FROM employees WHERE title='Senior Engineer' AND '1986-06-26' BETWEEN from_date AND to_date;`.

```
[mysql> SELECT COUNT(*) FROM titles WHERE ('1986-06-26' BETWEEN from_date and to_date) AND (title='Senior Engineer') LIMIT 10;
```

COUNT(*)
2795

```
1 row in set (0.62 sec)
```

3. There are more than 150,000 employees in our record have or once had the title containing "Engineer". Here I presented 10 randomly selected employees.

```
[mysql> SELECT COUNT(*) FROM employees WHERE emp_no IN (SELECT emp_no FROM titles WHERE title LIKE '%Engineer%');
```

COUNT(*)
150888

```
[mysql> SELECT emp_no, first_name, last_name FROM employees
[   -> WHERE emp_no IN (SELECT emp_no FROM titles WHERE title LIKE '%Engineer%')
[   -> ORDER BY RAND() LIMIT 10;
```

emp_no	first_name	last_name
208250	Arvind	Ellozy
210200	Kazuyasu	Dalton
488390	Alenka	Nilsson
440933	Xiadong	Kavanagh
233541	Amalendu	Ecklund
459569	Honglan	Schrift
242171	Claudi	Yurov
22006	Ortrun	Ladret
410224	Fun	Mellouli
449216	Berhard	Ratnakar

```
10 rows in set (1.42 sec)
```

The data dictionaries for the tables I used above are attached as a spreadsheet in the email. Let me know if you have any additional questions.

Best, Jonas